Improving and Assessing Students' Line Graph Interpretations: The Case of the Graph-As-Picture Interpretation

Grecia García García

Submitted for the degree of Doctor of Philosophy

University of Sussex

January 2015

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:

Grecia García García

A mis padres: Enriqueta y Rafael;

y a mis hermanas: Roma y Francia

Acknowledgements

First of all, I would like to thank my two supervisors (in alphabetical order): Prof Peter Cheng and Dr Richard Cox. Their guidance along all these years has been invaluable.

I would like to thank my sponsor in Mexico, CONACyT. Without its support, I would not have achieved this.

I would like to thank the teachers and staff from the primary schools that participated in my study. I want to express my gratitude to all my participants, who always made my day shine a bit more. I want thank to their parents too, who allowed me to spend time with their children.

I would also thank Dr Sharon Wood and Dr Judith Good for their comments and advice during the annual reviews. I would like to thank Dr Pablo Romero for the helpful discussions we had, and Prof Emmanuel Manalo for his comments, criticisms and reviews on several parts of this thesis. Also, I would like to thank Eric Harris for helping me to build the circuits I needed for the infra-red finger tracking device used in my experiments.

Over the years, I have enjoyed and learned from my colleagues and friends who I met at the Representation and Cognition Lab: Beate Grawemeyer, Ronald Grau, Alberto de Obeso, Andrés Gómez, Christopher Smith, Robin Kent, Erlijn van Genuchten, Unaizah Obaidellah, Maria Zulkifli, Norkhairolizah Hamzah, Noora Fetais and Asli Eyecioglu. Thank you very much for your support.

I would also like to thank the Mexican and Spanish families who I met in UK and have made me feel part of them: Héctor Rojas, Mariana Trejo, Alexis Rojas, Susana Fierro, Ashish Umre, Rohan Umre, Ferran Umre, Karla Pérez, Julián Salazar, Andrea Salazar, Mateo Salazar, Juan Loera, Ana Mendoza, Genner Llanes, Cynthia Marín, Francisco González (Curro), Carolina García, Diego García and Nicolás García.

I would like to thank my Aikido group, which has maintained my health in soul and body. Particularly, I would like to thank, my sensei, Jerome Chin Aleong, who has been very comprehensive and supportive all these years.

I want to thank my friends who have come and go, but that have left me great memories of our time together. Also, a huge thank you goes to Eric Tam, whose friendship has never failed me.

A very especial mention goes to Rossano Barone, who has been my colleague, friend and has been with me in good and in difficult times; I own him so much... including learning how to cook.

Finalmente, quiero agradecer a mi familia por dejarme sentir todo su cariño en todo momento, sin importar que un océano nos separe. Ellos siempre han creído en mi y han sido un gran ejemplo a seguir.

UNIVERSITY OF SUSSEX

Grecia García García, Doctor of Philosophy

Improving and Assessing Students' Line Graph Interpretations: The Case of the Graph-As-Picture Interpretation

Summary

The "graph-as-picture misconception" (GAPM) occurs when an abstract representation (e.g., a line graph) is interpreted as a picture of an object (e.g., a mountain). Previous research on students' line graph interpretations has focused on secondary school level and above, thus this research extends the investigation of the GAPM to primary school level. Particularly, it investigates: which type of environment is more effective for improving young students' line graph interpretations; and how can be assessed their interpretations.

A pilot study involved an improved version of Janvier's (1978) paper-and-pencil tasks (to create an interactive learning environment) and it investigated how to incorporate a card-sort task (to assess students' interpretations). Different touch-screen technologies were considered too.

Two experiments were conducted. In experiment one, 37 participants (third to sixth year) were assessed in their graphical knowledge through a picture/diagram card-sort task and a "pictorial group" was formed using participants' interpretations. During the intervention, students performed an active or passive mode of a Racing Car activity in which they moved or watched a car along a track while its speed/distance graph was plotted concurrently alongside. The results suggested that a wide variety of pictorial interpretations exist and students seemed to benefit from the active modality.

In experiment two, 38 fifth-year students performed different assessment tests. Extending experiment one, a "drawing the graph" mode and its passive modality were included. In that mode, students modified a plotted line of a speed/distance graph, which was used by the system to race a car along a track. Previous results were not confirmed: only students under the "drawing the graph" modality (including the "pictorial group") sig-

nificantly improved their interpretations; and different assessment tests seemed better to observe students' various interpretations.

In conclusion, a learning environment that allows interaction with the representation could potentially improve students' interpretations, which might be better assessed through a rich set of tests.

Contents

List of Tables

List of Figures

Publications

The following publications derived from parts of this work. Parts of Chapter 3 were published in:

Garcia Garcia G., Cox, R.: Children who Interpret Graphs as Pictures. In: V. Dimitrova, R. Mizoguchi, B. du Boulay, A.C. Graesser (eds.) Proceeding of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation To Affective Modelling. Vol. 200, pp. 695–697. IOS Press, The Netherlands (2009)

Garcia Garcia G., Cox, R.: An infra-red finger tracking system used in the assessment and remediation of "graph-as-picture" misconceptions. Proceedings of the Workshop on Technology-Enhanced Diagrams Research, Canterbury, UK, July 2nd, 2012, `ceur-ws.org/Vol-887/GarciaGarciaandCox.pdf`.

Parts of Chapter 4 were published in:

Garcia Garcia G., Cox, R.: "Graph-as-picture" Misconceptions in Young Students. In: A.K. Goel, M. Jamnik, N.H. Narayanan (eds.) Diagrams 2010. LNAI 6170, pp. 310–312. Springer, Heidelberg (2010)

Garcia Garcia G., Cox, R.: An Interactive Educational Diagrammatic System for Assessing and Remediating the Graph-as-Picture Misconception. In: V. Aleven, J. Kay, J. Mostow (eds.) Tenth International Conference on Intelligent Tutoring Systems: Bridges to Learning 2010. Vol. 6095, pp. 224–226. Springer, Heidelberg (2010)

Other publications include:

Garcia Garcia, G., Cox, R.: Diagrams in the UK National School Curriculum. In: G. Stapleton, J. Howse, J. Lee (eds.) Diagrams 2008. LNAI 5223, pp. 360–363. Springer, Heidelberg (2008)

Garcia Garcia G., Cox, R.: Does the K-12 practice of mixing pictures and graphs contribute to students' 'graph-as-picture' misconceptions? Abstract presented at the 31st Annual Conference of the Cognitive Science Society Amsterdam: Cognitive Science Society (2009)

Chapter 1

Introduction

This research is about young students' line graph comprehension and about how we can facilitate and assess their interpretations. We investigate which type of learning environment medium supports their interpretations of line graphs and how we can improve current methods to assess their graphical knowledge. Special attention is given to their pictorial-like interpretations as well. An overview of this work is presented in the next sections.

1.1 Motivation

Look at the item presented in Figure 1.1: What can you say about it?

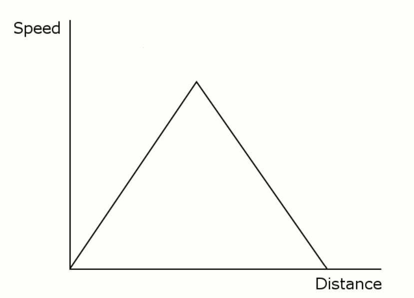


Figure 1.1: What can you say about this item?

An answer from one of our participants is presented below. Descriptions of the student's hand/finger gestures are underlined and within brackets:

> It's a shape of a mountain and again, if you are cycling, it will take you a long time {<u>he points along the first part of the plotted line</u>} to get up there because it's very steep and when you go down you can get on fast a bit {<u>he points along the second part of the plotted line</u>} (participant 21)

Responses like the one participant 21 gave are well-known to educators and researchers. In fact, this type of interpretation is known as the "graph-as-picture misconception" (see [42], [38], [10], [45]). That is, when a student interprets an abstract representation (e.g., a line graph) as a picture of an object (e.g., a mountain) it is said that the student is giving a graph-as-picture interpretation answer. Thus, in the example above, participant 21 manifested a pictorial conception of the graph since participant 21 was interpreting the plotted line as the slopes of a real mountain. Let's look at the interpretation given by another of our participants:

> You are going really fast {he points along the first part of the plotted line} and slow down {he points along the second part of the plotted line} and you will be doing quite a distance {he points along the x-axis} (participant 8)

In spite of the lack of a title label and scales in the axes, participant 8 provided a correct abstract interpretation of the graph.

Participant 21 and participant 8 are in the fifth primary (elementary) school year[1]. However, participant 21's young age does not necessarily justify his incorrect interpretation. Likewise, participant 8's ability to provide a correct answer should not be really surprising for his age.

Firstly, some research suggests that young students are able to understand external representations as line graphs or hierarchies (see [1], [19], [36]). Thus, participant 8's utterance should not be surprising. If there is some evidence suggesting that young students are able to comprehend some external representations (e.g., line graphs) then it is important to ask why is there not more research on how we can support young students' graphical competency? The lack of research is particularly surprising, especially when calls to improve students' graphical knowledge (graphicacy) have been made (e.g., [6], [2], [57], [72]) and its importance to help reasoning [23], support problem solving [69], enhance memory [8] or promote discovery [72] has been pointed out so many times.

Secondly, the kind of pictorial interpretation shown above has been observed in students at secondary school level and above. Furthermore, there are studies that show examples of teachers and professionals giving a graph-as-picture interpretation (e.g., [7], [70]). That is, there is evidence that iconic/pictorial conceptions occur across different levels in school and can be found even among professionals. In spite of the seemingly widespread occurrence of this type of interpretation, to this author's knowledge, little research has been done towards investigating young students' (pictorial) interpretation of

[1] In the United Kingdom, students' ages in the fifth school year range from 9 to 10 years old.

abstractions such as line graphs.

There is one more point to note about iconic/pictorial interpretations. The reader is invited to assess the following interpretations of the line graph presented in Figure 1.2:

"She cycled up a steep mountain"

"She kept speeding until she reached the top"

"She rode northwards and then went south"

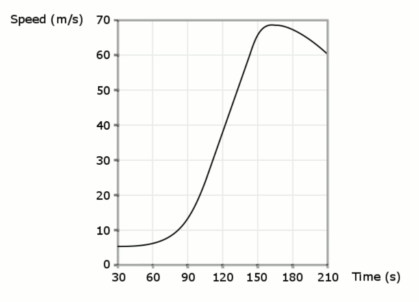


Figure 1.2: A line graph.

Under the traditional definition of the "graph-as-picture misconception", all those answers have been described as graph-as-picture interpretations. Nonetheless, such a term is rather too general to characterise all of these individuals' conceptions; that is, it disregards different pieces of pupils' knowledge. For example, the second interpretation seems closer to a correct interpretation than the others (*"She kept speeding ..."*), but this distinction is not made in the traditional definition. Overall, it seems that there is little research on the type of iconic/pictorial interpretations made by primary school students. Furthermore, current analysis and assessment of children's interpretations do not sufficiently consider the full richness of their often fragmented knowledge. This means that we may need to take another approach to evaluate students' knowledge of line graphs.

Thus, some of the questions that remain unanswered are: What kind of learning environment could support students' interpretations of line graphs? How can we assess students' line graph interpretations?

1.2 Aims

This thesis is mainly concerned with supporting primary school students' line graph interpretations and improving the way their interpretations are assessed – particularly for those students who give pictorial answers. Thus, we aim to answer:

(1) Which kind of environment could be more effective to improve students' line graph interpretations?

(2) How we can assess students' interpretations of line graphs?

That is, our objective is to investigate the kind of learning environment that could be more suitable for young students to learn about line graphs. Research on this area has somehow been inconsistent, while some pieces of work point out that active environments might be more effective than passive environments, others do not confirm that. Furthermore, we might need a better way to evaluate students' knowledge of line graphs since it is not clear whether students' knowledge has been accurately assessed (with regard to pictorial conceptions).

We explore these issues in what can be regarded as one of the first formal educational settings for graphicacy (primary school). In addition, we investigate these questions based on how we can use technology to enhance learning (learner-centred approach, [48], p. 9) and how technology can be used to support research (technology-enhanced research, [22]).

1.3 Overview of the Thesis

An overview of this work is presented.

Chapter 2: Literature Review

This chapter has two parts. The first part starts with a review of the research literature on students' knowlege, graphicacy and a brief overview of what research says about young students' knowledge of external representations. It continues with a review of students' difficulties when interpreting line graphs and it gives details of what is called the "graph-as-picture misconception".

The second part continues with a presentation of some research that studies how to improve students' graph comprehension through interactive learning environments. We review some of these studies since they were particularly useful in guiding the construction of our learning environment. In particular, Janvier's tasks [38] are described here since his research has been influential in this area. Janvier's tasks are used as a base to construct our batteries and main activities as well. Afterwards, we present some of the research that has been done to assess students' diagrammatic knowledge.

Chapter 3: Pilot Study and Technology Development

The first part of this chapter presents a pilot study of the interactive learning environment and assessment method used to support and observe students' line graph interpretations. Specifically, two of Janvier's [38] paper-and-pencil tasks (a Graph Comprehension Questionnaire and a Racing Car activity) were redesigned to be used with young children and were computerised to create an interactive learning environment. The tasks were evaluated through the experiences of a small group of young pupils. General observations pointed out that the proposed tasks were generally suitable for young students; however, some improvement was needed, particularly on the suggested method to assess students' graphicacy.

The second part of the chapter presents a summative usability study of three different interactive technologies: a NEC MultiSync LCD touchscreen, an HP TouchSmart, and a Wiimote[2]-based system. We discuss some of the problems and advantages of them as interactive learning technologies and we offer a recommendation on which of them seems more adequate for our experiments.

Chapter 4: Experiment 1

In this first experiment, the tasks from the pilot study were improved and administered to a group of primary school students. Students were assigned to either the active or the passive mode of a Racing Car activity. Additionally, an improved graphical discrimination task was used to observe pupils' graphical knowledge.

Results suggested that the active modality was more useful for improving students' line graph interpretations than the passive modality. However, (1) this result needed to be investigated further due to the observed ceiling scores in the assessment task. We also found evidence of various types of pictorial interpretations; thus, (2) a better assessment test, which could consider those varieties, was needed. In addition, students' behaviours during the Racing Car activity revealed a "drawing the graph" strategy. Hence, (3) the design of a new active environment that takes into account this strategy was suggested. Those three relevant points were explored further in experiment 2.

Chapter 5: Experiment 2

In the second experiment we focused on confirming the results from experiment 1, improving the assessment test and extending the active modalities used in the previous

[2] http://johnnylee.net/projects/wii/

experiment (by adding an active mode that allowed changing the graph).

Although we could not confirm the results from the previous experiment, other important outcomes were obtained: our new active modality (based on the “drawing graph” strategy) seemed promising for improving students’ line graph interpretations and therefore reducing pictorial interpretations. Again, the students’ answers observed varied largely and, as a result, we noted that it might be better to assess their interpretations using a rich set of tests. That is, the way students’ graphical knowledge is assessed could be improved by using different assessment methods so that students’ knowledge and beliefs are more explicit to, for instance, educators.

Chapter 6: Discussion and conclusions

This chapter presents a general discussion of the results obtained in the empirical work. It reviews our research questions, how we tackled them and what our outputs were. For each question we offer an explanation of our results. Alongside this, we explain some of the issues that might arise from our approach and some of its advantages too. We end with a summary where (1) we point out that the “drawing the graph” type of interaction is promising to improve pupils’ line graph interpretations and (2) we support assessing pupils’ pieces of knowledge through a rich set of batteries.

Chapter 2

Literature Review

2.1 Introduction

This chapter starts with an overview of what we understand for students' knowledge (Section 2.2) and particularly for students' knowledge of diagrams and graphs (Section 2.3). This is followed by a review of some of the most prominent students' difficulties when interpreting line graphs (Section 2.4), where we pay particular attention to the one known as the "graph-as-picture misconception".

In the next part (Section 2.5), we revise how learning environments could improve students' knowledge of a variety of complex subjects such as physics, mathematics and, to a certain extent, graphicacy. However, we note that more research needs to be done regarding the design of diagrammatic learning environments in order to address some of the mixed results found in the literature about their effectiveness.

Afterwards, we discuss some methods used in research to assess students' knowledge of graphs, but we point out that current assessment methods might not be accurate in describing students' knowledge state (Section 2.6). In order to support this point, we provide a few examples of different students' interpretations that have been loosely characterised as "graph-as-picture misconceptions".

At the end of the chapter (Section 2.7), we lay out two issues that we address in this research: (1) finding out what type of environment could be more effective to improve students' knowledge of line graphs and (2) how to improve methods to assess students' knowledge of line graphs.

2.2 Students' knowledge

The problems students exhibit during instruction have been described in different ways. For example, students' conceptions have been described according to their compatibility with accepted meanings (terms such as *misconceptions* [17, 45, 53] and *alternative conceptions* [52] are an example of this), or whether students' inadequate conceptions happen before or after instruction (e.g., *preconceptions* [17]), or more traditionally they have been considered as *errors*.

These definitions suggest that an implicit dichotomy exists between what it is considered "accepted knowledge" and what is not. For instance, Mevarech and Kramarsky made a distinction between knowledge that is compatible with accepted meanings and knowledge that "*differs from that which is to be learned*" ([52], p. 231); and Clement, Brown and Zietsman [17] distinguished between students' ideas that are incompatible with accepted scientific knowledge (*misconceptions*) and compatible (*anchoring conceptions*).

Nonetheless, these definitions seem poor when describing students' problems. That is, these definitions rigidly classify students' knowledge. According to Leinhardt et al.:

> ...students' knowledge... arise largely from everyday experience... may involve a mixture of everyday and deeply understood formal knowledge... ([45], p. 24).

In other words, a classification of students' knowledge as "acceptable" or "unacceptable" does not seem appropriate. Mokros and Tinker made this clear when reporting their participants' ideas:

> there is a hierarchy of misunderstandings and incomplete ideas which a student brings to instruction at any point, all of which impede, to a greater or lesser extent, a student's progress... some of the errors can be very idiosyncratic and difficult to categorize ([53], p. 380).

This indicates that even classifying students' problems could be difficult to achieve. That is, a categorical definition of students' difficulties does not seem adequate since knowledge is complex and intertwined with day-to-day and formal concepts.

Furthermore, Smith, diSessa and Roschelle [64] stated that the variation between terms to describe students' difficulties reflects *"how researchers have characterized the cognitive properties of student ideas and their relation to expert concepts"* (p. 119). For example:

> The qualifiers, *pre*, *mis* and *alternative*, each point to different presumptions about the nature and origin of conceptions. Likewise, *belief* and *conception*

> suggest unitary cognitive structures, whereas *theory* and *framework* embed conceptions in larger scale structures ([64], p. 119).

Regarding this, Smith et al. [64] proposed a characterisation of students' knowledge in terms of novice and expert knowledge:

> Where misconceptions research has focused on the discontinuities between novice students and experts, we identify and emphasize important dimensions of continuity between them (p. 116).

That is, knowledge could be described as a transition towards expertise:

> student's continual reconstruction of their existing knowledge may produce the observed intermediate states of understanding and eventual mastery of a domain ([64], p. 117)

Considering this perspective, some of the well-known students' difficulties (Section 2.4.1) would be happening along that transition. Assuming that knowledge would broadly differ between people and it could be intertwined with accepted concepts (ideas), those common difficulties might not be displayed/exhibited in the same way from person to person.

More formally, diSessa explains the transition from novice to expert knowledge using a knowledge system named "knowledge-in-pieces" [28, 64]. In this mechanism, diSessa explains how simple abstractions obtained from experience (*phenomenological primitives* or *p-prims*) become complex knowledge structures. P-prims are small knowledge structures that are self-explanatory, are activated (cued) according to context and they originate in superficial interpretations of experienced reality [28]. Expertise is achieved by adding new p-prims, reorganising/enhancing/reducing their inter-relationships and changing function (e.g., as heuristic cues) ([28], p. 114-115).

According to this view, a more appropriate criterion to characterise triggered knowledge structures would be to describe them as productive or unproductive. For example, when a student conception leads to an erroneous conclusion in one context then it is said that the particular knowledge structure is unproductive in that specific context. Thus,

> commonly reported misconceptions represent knowledge that is functional but has been extended beyond its productive range of application ([64], p. 152).

In other words:

> misconceptions... have their roots in productive and effective knowledge. The key is context – where and how those conceptions are used ([64], p. 124-125).

Therefore, following diSessa and colleagues' framework, we will consider that students' knowledge is formed by a mixture of novice and expert conceptions which, when applied in a particular context, could be productive or unproductive. From now on, when talking of "'misconceptions", "problems", or "errors", we will be using these terms within this definition.

In this thesis, in particular, we are interested to study students' knowledge of external representations (see next section). Thus, in the next section we give an overview of the research that has been done in this area before presenting some of the difficulties students face when working with a specific external representation (i.e., line graphs).

2.3 Graphicacy

Cox and Brna [23] stated that:

> The term 'external representation' (ER) refers to a wide variety of representations in both the linguistic and graphical modalities (p. 242)

However, in the work presented here, the term *external representation* (ER) includes only graphical or analogical representations (e.g., bar graphs, maps, set diagrams) and intermediate forms between those modalities (e.g., tables, lists). Pure linguistic representations (such as sentences of natural language or formal language) are excluded for the purposes of this work.

External representations are important because they could support reasoning [23, 20], facilitate problem solving [69], enhance memory [8], support communication [21, 69], assist learning [3], and promote discovery [72]. Thus, it is not surprising that not only calls for integrating them in education have been made for decades, but a recognition of a fourth academic skill in education has been required: *graphicacy*.

Aldrich and Sheppard [2] defined graphicacy as:

> the ability to understand and present information in the form of sketches, photographs, diagrams, maps, plans, charts, graphs and other non-textual, two dimensional formats (p. 8).

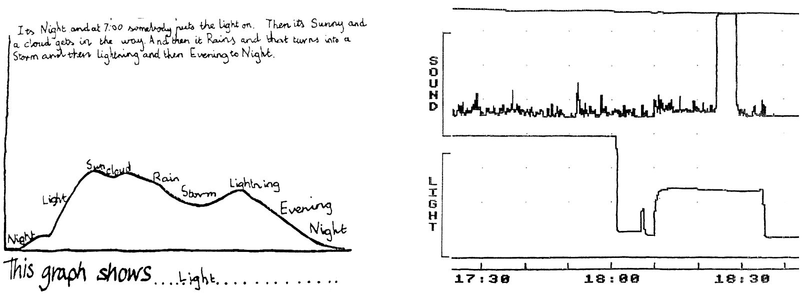
Balchin and Coleman [6] noted that graphicacy is inter-disciplinary, it *"spills over into literacy ... and numeracy without being more than marginally absorbed in either"* (p. 947) and it complements words and numbers but does not replace them. That is, graphicacy emphasises and promotes the idea that external representations are a multifunctional tool *to think with* – as Tolchinsky ([67], p. 2) put it.

Graphicacy is not usually taught in the curriculum and children are expected to "pick it up" as they go along ([2], p. 8). Phillips [58] added that it is largely limited to the use of bar graphs and maps in primary school level (p. 50) and Garcia Garcia and Cox [33] showed that this is still the case. In addition, although technology has facilitated the construction of a variety of ERs, the end result is not always meaningful [58]. This problem is aggravated when teachers do not use appropriate representations during lessons [68], teaching activities foster "algorithmic graphing" [57], or teaching resources (such as text-books) present substantial inconsistencies [2]; all these cases undermine the learning process. As Wainer [72] put it, *"even though ... the ability to understand graphically presented material is hard-wired in ... the ability to draw graphs well is not. It requires instruction... "* (p. 18).

Although the curriculum is already packed, there are plenty of opportunities to include graphicacy. For example, Aldrich and Sheppard [2] spoke of CLEAR, a five-step criteria for using educational graphics. It encourages educators to ask themselves whether the graphic is conveying its **c**entral point, to observe if the **l**ayout is well-designed, to note if the graphic is a good **e**xample of its type, to consider what prior knowledge is **a**ssumed by the learner and, if the representation is critical, to **r**edesign it themselves ([2], p. 10-11).

Pereira-Mendoza [57] advocated targeting three specific aspects of graphicacy instead: exploring assumptions of the data, discussing alternative representations, and predicting from the data. Pereira-Mendoza [57] exemplified this within the context of "the number of pets that children have" in order to (1) consider data in different forms (e.g., how to count "lots of fish in a tank": Do we count the number of fish or do we count the number of tanks?), (2) to represent data in different formats before deciding which is clearer (e.g., grouping data in columns using even icons or using different icons), and (3) to provide an opportunity to predict from the diagram (e.g., which pet a child is more likely to have?).

This thesis is a contribution to improve students' graphical skills, but our focus is on line graphs only. Thus, the next section presents some of the research that has been done specifically in this area.



(a) Line graph drawn by a student.

(b) Line graph produced by `The Measuring Box`.

Figure 2.1: Examples of the line graphs constructed and interpreted by primary school students. Source: Both images from [54] (p. 11-12).

2.3.1 Line graphs

Although line graphs are one of the most commonly used ERs, a large part of the literature on this topic focuses on secondary or tertiary school students (e.g., [45], [9], [53], [38]). Thus, in order to cover part of the missing literature, we direct our research to the study of young students' understanding of line graphs. Then, this section presents a review of primary (elementary) school students' line graph abilities.

Morgan [54, 55] and Corcoran [19] reported that young students are able to understand and draw their own line graphs. Morgan [54], in particular, gave a detailed explanation of her experience with a class of 7- and 8-year-olds. The class was divided in two groups. One group was given the task to graph a story, which included changes of light, temperature and sound. The other group was introduced to `The Measuring Box` (a computer-based data logging system which records sound, light, movement and temperature levels) and children were asked to interpret the line graph produced by the `Measuring Box`, which was left overnight at the school recording sound levels. On the one hand, Morgan [54] observed that students who draw their own graph provided accurate representations of their stories, showing high levels of understanding. One example of a student's graph is in Figure 2.1a: details of the situation (e.g., passing clouds and a storm) were drawn with accuracy. On the other hand, students who used the computer exhibited progression from an interpretation task to an enquiry task. For example, for the line graph in Figure 2.1b, students first observed (interpreted) a loud sound around 18:30 hrs. and then they were keen to investigate the cause of that loud sound recorded by the `Measuring Box`. Although students showed some competency when interpreting line graphs, Morgan [54] noted that children needed time and more experience to consolidate what they learned before moving on to more complex line graphs.

Morgan [54] and Corcoran [19] did not report to what degree their groups were successful to understand or create their line graphs. Nevertheless, in another and more extensive project, Ainley [1] reported that more than half of a class of fifth year students (9-10 year-olds) were able to construct their own line graphs and to interpolate from them without explicit instruction. For example, in Ainley's study, students were given a table with data of a child's height at different ages and students were asked to plot that data on a computer. Pupils were later asked to read the age or the height of the child in points that were not marked on the graph. A parallel activity was done for another group of students who received the tables and graphs on paper. Although Ainley did not find differences in the performance between those two groups, she claimed that students performed better than was usually expected from primary school students not because they were *not* taught how to use line graphs, but because the task was meaningful to them:

> In contrast to the way in which line graphs are traditionally introduced, we did not deconstruct the task into constituent skills, but presented the children with complete images of the graphs in contexts which focused not on the representations themselves, but on the meaningful tasks using the graphs ([1], p. 259).

Ainley [1] also noted that the graphs presented to pupils had two characteristics. First, the data presented was artificial, but meaningful to students. Second, the pattern of the plotted line matched the phenomenon (e.g., increasing height is reflected in an increasing trend in the plotted line).

Nonetheless, research by Beveridge, Dean, Avons and Hitch [13] showed that 10-11 year-olds are capable of understanding journey graphs (time-distance graphs); that is, they presented their students with data where the plotted line does not necessarily match with the phenomenon. Beveridge et al. [13] tested three procedures to teach journey graphs. The "concrete embedded" procedure simulated the situation using a concrete reference where a paper cutout of a boy was moved over a map and its journey graph was plotted. The "hypothetical embedded" procedure gave the students a verbal description of the situation only. The "disembedded" method gave no reference to the background story to the students. Results of this study (which included interpreting graphs, extracting information from the graphs and graph construction) showed no significant teaching procedure effect, but authors claimed that *"children in this range age are capable... of making considerable progress towards understanding journey graphs"* ([13], p. 152). They concluded that *"children who can be presumed to understand concepts of time, distance and speed are capable of learning to represent relationships between these concepts using*

Cartesian coordinate graphs" ([13], p. 153).

Overall:

1. Children in the third and above school years might be capable of interpreting and constructing simple line graphs; that is, graphs which show one plotted line and use everyday concepts such as time, distance, noise, temperature, light and speed.

2. Technology provided an opportunity for enquiry and it could motivate students during the tasks (as was the case with Ainley [1], Morgan [54] and Corcoran's [19] activities).

In addition, note that in some research, like the ones mentioned above, assessment of students' abilities was performed using methods such as: graph construction [54, 13], multiple-choice instruments [26], group discussions [1], and interviews [1, 26]. We use some of these methods to observe students' interpretations, but we will also try to use an alternative method to get an insight into students' graphical knowledge (see Section 3.1.2, Section 4.3.1 and Section 5.3.1).

In this section, we stated the importance of graphicacy and presented some evidence of young students' abilities to interpret line graphs, but students might face some difficulties when using ERs too. In the next section, we indicate some of the difficulties students face, but we will focus on students' problems when interpreting line graphs.

2.4 A brief summary of students' difficulties when understanding line graphs

In the previous section we showed that primary school children are capable of using line graphs (Section 2.3.1). However, little research has been done regarding primary school students' difficulties. Thus, in the following subsection, we briefly mention some of the well-known problems students face when interpreting line graphs. In particular, we introduce details of one of the well-known difficulties known as the "graph-as-picture misconception" and we put it in context of diSessa' framework [28, 64], which we discussed in Section 2.2.

2.4.1 Students' difficulties when interpreting line graphs

There are several difficulties that students can face when interpreting line graphs, and a very brief summary of some of them is presented here.

Bell, Brekke and Swan [10], McDermott, Rosenquist and vanZee [50] and Leinhardt, et al. [45] described some of the more common students' difficulties as related to:

- Problems relating to two variables,
- Tendency to look for linear patterns,
- Interpreting intervals point-wise,
- Confounding the slope and the height of the graph, and
- Seeing a graph as a picture.

Bell et al. [10] described some students who could look at one variable at a time on a graph and they were not able to observe the relationship between both variables. For example, in a cost/weight graph of sugar prices two different students gave readings of either the values of the cost axis or the weight axis, but they were not able to relate the change of the price with respect to the weight.

Another problem that Leinhardt et al. [45] and Bell et al. [10] described is the tendency to look for linear patterns. For example, some students could find difficult to see the "curved variations" between the variables; thus, they use the graph as if it were constituted by linear segments. Bell et al. [10] claimed that students who see the graph point-wise are prone to this form of interpretation.

Leinhardt et al. [45] described interval/point confusion as the student focusing on one point rather than a range of points. For example, students could respond with a single point when asked when the sound was really loud in Figure 2.1b instead of giving an interval of time.

McDermott et al. [50] and Leinhardt et al. [45] described slope/height difficulty as students confusing two different features of the graph. McDermott et al. [50] stated that, when interpreting *"a straight-line graph, for example, information may be contained in the coordinates of a point, the difference in the coordinates of two points, ... or the slope of a line"* (p. 504); thus a slope/height difficulty happens when a person incorrectly chooses the information given by a point instead of by the slope.

Kerslake [42], Leinhardt et al. [45] and McDermott et al. [50] provided more examples of many other students' problems. The reader interested in the topic is encouraged to look at that research. The description regarding "seeing the graph as a picture" is described in the next section.

2.4.2 The "graph-as-picture misconception"

One of the common difficulties when interpreting line graphs is known as the "graph-as-picture misconception" [10, 16, 7]. It happens when an abstract representation is interpreted as a *picture* of a situation [10]; that is, *"confusing a graph of an event with a picture of the event"* ([7], p. 2). For example, a line graph in Figure 2.2 could be interpreted as somebody going "up and down a hill and up again".

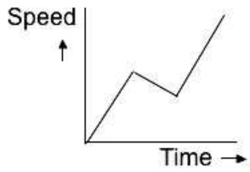


Figure 2.2: A line graph.

Janvier [39], who was one of the first researchers to look at the graph-as-picture interpretation (GAPI) in more detail, described it as *"the simultaneous use of the graph at the symbolic and pictorial levels"* (p. 118). In addition, he [39] also observed that the GAPI could happen even after the students have been able to extract data from the graph (e.g., after identifying the maximum speed of the graph in Figure 2.2, students could interpret the graph as a hill).

In terms of the knowledge-in-pieces framework, Elby [29] suggested that a GAPI is related to a p-prim he called *what-you-see-is-what-you-get* (WYSIWYG), which means that a representation has been interpreted iconically or realistically (p. 486). As we mentioned before, p-prims are defined as simple abstractions from common experiences, but they are also triggered by context. Examples of when WYSIWYG is critically useful are: a sign on a road illustrating a curve in the road ahead or a label of a flammable substance.

Since applying WYSIWYG can be productive in a variety of contexts, it could get cued strongly. Elby [29] also proposed that a compelling visual attribute might trigger the WYSIWYG structure, *"edges, corners, and motion probably constitute compelling visual attributes..."* (p. 488). He suggested that:

> most useful visual attributes become involved in quick and direct interpretive strategies... By their very nature these quick-and-direct strategies disproportionately grab attention ([29], p. 490).

Overall:

> WYSIWYG is one of the intuitive knowledge elements contributing to a "naive" interpretation of a visual representation... WYSIWYG can contribute to both productive... and flawed... interpretations ([29], p. 486).

Although Elby pointed out that his suggestion needs to be backed up by more empirical evidence, Elby's interpretation of GAPIs in terms of knowledge-in-pieces is roughly in agreement with Clement's explanation of the same difficulty, our own findings (presented in Chapters 4 and 5) and reported students' utterances from other research (see section 2.6.3). More specifically, Clement noted that a student could be:

> making a figurative correspondence between the shape of the graph and some visual characteristics of the problem scene ([16], p. 372).

Clement noted two patterns on students' pictorial interpretations, which agree with Elby's suggestion of the WYSIWYG being triggered by compelling visual attributes. Clement [16] named those patterns as "global correspondence error" and "feature correspondence error", which relate to matching the *"entire problem scene... to the shape of the entire graph"* and matching *"a specific visual feature of the problem scene... to a specific feature of the graph..."* ([16], p. 374), respectively.

Thus, we will consider Elby's account of GAPIs along with this work. Before continuing with a review of the literature on what has been done to improve students' understanding of graphs (and which will take us to our first research question), we need to point out why we give attention to students' graph-as-picture interpretations:

- It is one of the most well-known forms of answers given by students.
- We have a possible account of its nature from other research (e.g., [29], [16]).
- Research literature on the "graph-as-picture misconception" offers us a basis on which we can build our research upon (e.g., Janvier's [39] tasks that help us with the design of a learning environment, and Rogers et al. [60] informed us of which characteristics we could implement in our design, etc.; see Section 2.5).
- By observing the rate of change of students' GAPIs, we could assess whether our learning environment has been effective and whether students have gained experience for understanding line graphs. In Elby's words:

    > compelling visual attributes end up with strong connections to WYSIWYG... [but] WYSIWYG gets cued less strongly and less frequently as

> students *gain experience* with abstract representations ([29], p. 500, italics added).

2.5 Improving students' graphicacy through technology

In this section we explain what types of learning environments (LEs) have been used to improve students' learning and their advantages versus traditional paper and pencil methods. We point out that results in this area are mixed; therefore, we still need to investigate whether an active or a passive LE could be more effective in improving students' knowledge of graphs.

2.5.1 Using technology to enhance learning and research

There is no doubt that technology has given us an opportunity to explore new ways of teaching and learning. In fact, every time some type of revolutionary technology has come to the market (e.g., radio, TV, PCs, tablet PCs, telephones) there is always some eagerness to revolutionise school or to solve many teaching and learning issues with it (e.g., [56], [25], [40]).

After many years of observing how those technologies have failed to solve the problems faced in school, a more critical perspective of the role of technology in education emerged: a learner-centred approach of software design started to be recognised over the technology-centred approach that was so predominant in the last century [48]. That is, with a learner-centred approach we aimed to adapt technology to the needs of the students based on how human cognition can be enhanced rather than students adapting to whatever current state-of-the-art technology is in vogue.

In parallel, we adopted a technology-enhanced research approach [22] and took advantage of the technology to support our research. This point was particularly emphasised by Janvier's relevant piece of work. Janvier [38] interviewed students who performed several graphical interpretation tasks using paper-and-pencil materials. During the analysis phase, he noted that video-recording could have provided a better insight into students' strategies and he even suggested using *"a more sophisticated technology, allowing the researchers to watch the successions of points scanned by the subjects on the graph"* ([38], p. 4.3). We did not go as far as using eye-tracking technology, but we video recorded and data-logged interactions.

2.5.2 Microcomputer-based labs

Literature shows that microcomputer-based labs (MBLs) have been tested regarding their effectiveness to improve students' learning of abstract representations. MBLs allow a collection of physical phenomena through probes, which are connected to a (micro)computer to present the data as it is being collected ([7], p. 3). That is, MBLs provide an alternative form of (re)presenting information to the learner that cannot be conveyed through traditional paper-and-pencil methods.

Some research that is relevant to this work either test how MBLs support students' learning or suggest improving their tasks using MBLs. For example, Avons, Beveridge, Hickman and Hitch [5] aimed to improve students' line graph interpretations by using an external control box that allowed students to race a computerised car at specific velocities while observing the car's speed on a line graph; Morgan [54] explored how pupils interpreted line graphs when using the `Measuring Box` (see Section 2.3.1) to continuously record sound levels, which were plotted onto a line graph; and Janvier considered that his paper-and-pencil tasks could be improved by using interactive tools that involve *"sensory-motor responses. Easy devices... which would enable a pupil to show... motions depicted by graphs..."* ([38], p. 3.13).

Note that one of the properties that is shared across MBLs is that they allow the user to freely manipulate the state of an object or event. We will use this characteristic to establish a distinction between environments that allow the learner to manipulate the system (we will call them "active" environments) versus environments that do now allow this (we will call them "passive" environments). That is, an "active" environment allows the learner to explore certain aspects of a system at will, whereas a "passive" environment does not.

It is important to note that although our definition of "active" and "passive" environments is based on a visible interaction between the learner with the technology, it does not necessarily imply the degree to which the user is cognitively engaged in the task. Regarding this, Chi [15] explains that *"undertaking an output-producing type of... activities does not guarantee... the generation of new information [by the learner]"* (p. 79) and, vice versa. *"Even though the majority of the studies rely on externalized outputs, the outputs can be internal ones"* (p. 79).

An important issue to mention is that, even if there is some enthusiasm to support the use of active environments, it is not always clear if they are more (or equally) effective as passive environments. This issue is discussed next.

2.5.3 Active vs. passive learning environments

In this section we present some of the research that supports the use of active environments (see Section 2.5.2) and we question the evidence that they present as well.

Beichner [9] explored which aspects of MBL are critical to help students improve their interpretation of graphs and he concluded that:

> feedback appeals to the visual and kinesthetic senses. A simple visual juxtaposition of event images and graphs is not as good as seeing (and "feeling") the actual event while graphs are being made ([9], p. 812).

That is he emphasised the importance of coupling a physical event with its corresponding symbolic representation. Beichner [9] also suggested that MBL treatments should be useful at reducing pictorial-like interpretations. He argued that since working memory has limited capacity:

> the simultaneous presentation of event and graph "makes the most" of the cognitive facilities available. This should make it easier to transfer the event-graph unit (already linked together) into long-term memory as a single entity ([9], p. 804).

Anastopoulou, Sharples and Baber [3] also investigated how the mapping of the learners' own body with a representation encourages reflection. In their study, one group of students were given hand sensors and their movements were used to create a graph that was continuously plotted. The other group observed the experimenter creating the graph instead of creating it themselves. Anastopoulou et al. [3] found that the group who created their own graphs performed consistently better than their counterparts. They argued that *"Physical movement can support the creating and testing of hypotheses"* ([3], p. 268); and they added that allowing direct manipulation gives a *"sense of personal control and engagement"* and provides *"continual coordination of movement with the symbols that the movement creates"* ([3], p. 269).

In Mokros and Tinker's [53] study, students used their own motions as well as a toy car to explore different graphs. They, as other researchers, observed the benefits of using MBLs. Mokros and Tinker [53] suggested that MBL reinforces many learning modalities (e.g., kinaesthetic, visual, aural) and it provides a scientific experience to analyse data and eliminates the "drudgery" of graph production (p. 381-382).

The results of that research suggest that MBL could support students learning of abstract representations. However, it is not clear whether the direct interaction with the system was beneficial. For example, Beichner [9] did not find significant differences between his groups, although the MBL-group performed better. Anastopoulou et al.'s [3] groups differed on the types of experiences they had (one group benefitted from unsuccessful attempts to solve the task whereas the other experienced only successful treatments). Regarding Mokros and Tinker's study, we noted that even though there was an overall improvement on students' scores on graph interpretations, one third of students "persisted" on interpreting a speed/time graph as a picture; in other words, a large number of the students did not seem to improve/change the way they interpreted the graph.

Not all of the research shows results in favour of an active modality. For instance, Keehner, Hegarty, Cohen, Khooshabeh and Montello's study [41] allowed a group of participants to manipulate a computerised 3D object and asked them to make inferences on how cross-sections of the object would look like. In a yoked design, another group of participants observed the manipulations made by their peers and were asked to solve the same tasks. Keehner et al.'s analyses revealed that participants in the passive condition performed as well as their counterparts when the manipulations observed corresponded to effective manipulations of the 3D object.

Thus, given the mixed results observed in the literature, we aim to investigate which type of environment would be more effective in improving students' graphical interpretations. In order to construct a research-based learning environment, we review some of the literature on key issues to consider when constructing learning environments next.

2.5.4 Effective learning environments

Scaife and Rogers [63] proposed a theoretical framework that informs us of some key design issues to consider when designing tools to improve the use of diagrammatic representations. They suggested:

- Facilitating inference by directing attention to key components of the task.
- Considering the trade-off between different mediums[1].
- Facilitating or providing opportunities to produce graphical representations.
- Considering what combination of ERs could be more effective[2].

Scaife and Rogers' [63] observations were based on their External Cognition framework, which is an analytic framework used to explain:

[1] For example, allowing the user to leave cognitive traces (e.g., making notes on a textbook) vs. providing feedback (e.g., providing feedback to new configurations of a piece of software) [63].

[2] For example, using static diagrams with text or animations with narratives.

> the *cognitive processing* involved when interacting with graphical representations, the *properties of the internal and external structures* and the *cognitive benefits* of different graphical representations ([63], p. 188).

That is, it considers (1) how different representations could reduce/augment the computational offloading for the task; (2) how the structure of a representation could make the problem easier or more difficult; and (3) how graphical elements could constrain the inferences made about the represented world ([63], p. 189).

Examples of the effectivity of the aspects mentioned above can be found in the literature too. For instance, Rogers, Scaife, Aldrich and Price [60] successfully tested this framework using a diagrammatic learning environment called `PondWorld`. They particularly focused on complex subjects where the use of abstractions was needed and where children tend to have difficulties reasoning about the underlying topic ([60], p. 2). Rogers et al. used "dynalinking" (see Section 2.5.5), which worked as feedback to the learner to explore mappings and made explicit relations/constraints between the external representation and the represented world.

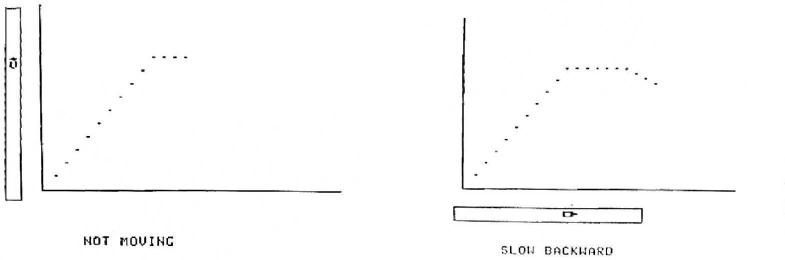
Another example of the effectiveness of the aspects mentioned above can be seen in White's `ThinkerTools` [73]. `ThinkerTools` present different graphical representations of force and motion (re-representation) to facilitate students' formulation of physics principles. The software allowed us to observe how changes in one representation affected the other (computational offloading). In addition, the simulation had different stages, which restricted the manipulations that students could make on the environment (constraint inferences).

Like this research, we will consider these design principles for the construction of our learning tool.

2.5.5 Supporting literature on learning environments

Other research is reviewed in this section since it informs us of additional characteristics we should take into account when building our learning environment.

The first piece of work is by Avons, et al. [5]. They introduced 9 and 11 year olds to a microcomputer-controlled display where they could move a car on a track and observe simultaneously the changes on a line graph. The car was moved using a "control box", which permitted the car to move (fast/slow) backward, not moving it, and (fast/slow) forward. Avons, et al. [5] tested two main dimensions: (a) the position of the track (along the vertical or horizontal axis); and (b) the type of interaction – either the student con-



(a) The track is parallel to the y-axis.

(b) The track is parallel to the x-axis.

Figure 2.3: Simulations used in [5]: The car and track are along the y-axis (left) and along the x-axis (right). Source: [5] (p. 97).

trolled the environment or the student watched somebody else controlling the environment. However, note that in the active condition the experimenter told students the sequence of movements they had to follow. Figure 2.3 shows the two different interfaces used in Avons, et al.'s experiment: a track and a car are parallel to the y-axis of a speed/distance graph (Figure 2.3a) or they are parallel to the x-axis of the graph (Figure 2.3b). A text with a description of the car's movement appears under the graph as well.

They [5] did not observe significant differences on students' performance by type of interaction. But importantly, they found that the position of the track seemed to affect students' comprehension. That is *"compatibility between the directions of movement in the simulation and graph may influence the ease which these associations are learned"* (p. 103, [5]). In this case, students using the vertical alignment performed better than students who had the horizontal alignment. However, Avons, et al. also concluded from their experiment that:

> two basic reasons why deeper understanding was not fostered... first... the presentation itself was inadequate. Possible shortcomings are: (a) An insufficient number or range of examples... (b) Inadequate interest and motivation... (c) Level of difficulty... displays of this type are too complex... A second reason may be that the children... were too young ([5], p. 104).

As we have discussed before (Section 2.3.1), young students are able to work with graphs as the ones shown in Figure 2.2 and Figure 2.3 (e.g., [19], [1]), therefore we could conclude that the representation they used was inadequate and we should consider this in this piece of work.

Research performed by Rogers, Scaife, Aldrich and Price [60] is relevant for this study too. The environment they designed, `PondWorld`, aimed to facilitate students' learning of ecological concepts. In particular, `PondWorld` focused on helping students to understand food webs. Their system presented students with a representation of a pond alongside a representation of a food web. `PondWorld` had several modules, which differed in the amount of cognitive work the student had to take. Importantly, it also provided *dynalinking*, which allowed the two representations to co-vary. In particular, Rogers et al. noted that:

> One of the main cognitive benefits of dynalinking is to allow relationships between elements of a complex concept(s) to be dynamically and explicitly displayed, together with the possibility of conveying mismatches and other conflicts between them ([60], p. 10-11).

They added that forms of computational offloading (as dynalinking) could act as a learning aid to understanding unfamiliar formalisms since it allowed students to understand mappings of multiple representations ([60], p. 29).

In parallel, Vogel, Girwidz and Engel's [70] observed the benefits of using dynalinking as well. Their research aimed to help students to *"bridge the gap between a concrete problem and its abstract representation"* ([70], p. 1288). They designed an environment that could explicitly present the link between two representations: the situation and the abstraction. In other words, they wanted to explicitly present *"operations which cannot be performed internally by the individual"* ([70], p. 1290); they called this "supplantation". Supplantation was achieved by using *dynamic linking*; that is, when one representation is changed, its corresponding representation changes simultaneously (p. 1290). For example, Vogel et al. considered that mental operations can be facilitated for graph interpretation when considering how *"a point on the graph is related to a concrete situation or object"* and how *"such a point and the concrete situation... co-vary"* ([70], p. 1290). A screenshot of Vogel et al.'s computer program is in Figure 2.4: it shows the relation between the radius and surface area of a cylinder when the volume is constant. Vogel et al.'s program [70] shows how a point on the graph is related to the concrete object and how the object changes when the point changes too. Overall, Vogel et al. noted that when solving different geometrical problems like this one, students' performance improved, even with lower achievers.

Note that some of the learning environments that support dynalinking could provide interaction across its representations (abstract or concrete). For example, one of the modules of `PondWorld` allowed the students to annotate the diagram only, whereas in another module the students could only interact with the pond [60]. Other environments that implement multiple dynalinking are `Graphs and Tracks` and `SimCalc MathWorlds.`®

David Trowbridge's `Graphs and Tracks` [49] (Figure 2.5a) is a tutoring software that

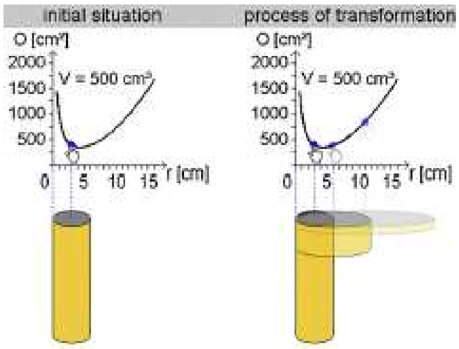
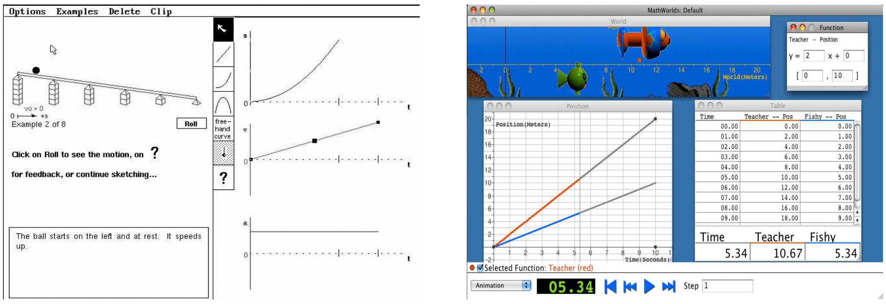


Figure 2.4: Screenshot of Vogel et al.'s application. Source: [70] (p. 1291).



(a) `Graphs and Tracks`: Part II – Motion to Graphs screenshot.

(b) `SimCalc MathWorlds`® screenshot.

Figure 2.5: Screenshot of two learning environments. The picture on the left was retrieved from [31].

helps students relate kinematics concepts with their graphical representations. It consists of two parts. The "Graphs to Motion" part asks students to make the appropriate modifications to a track in order to match the motion of a ball (that runs on the track) with the motion described on a given graph. The "Motion to Graphs" part asks the student to draw the motion graphs after observing how a ball rolls on the presented track. This software was based on research about students' understanding of velocity and speed concepts and on research about students' difficulties when interpreting motion graphs [49, 50]. `Graphs and Tracks` was used as an instructional and research tool [49, 35]. In particular, it was found to be useful for investigating students' thinking [35] and as a tutoring tool to *"relate the graph of a motion to the actual motion"* of an object ([49], p. 453).

`SimCalc MathWorlds`®, from Kaput Center[3], is an interactive learning environment designed to help students to understand algebraic concepts. Besides permitting students to manipulate complex graphs, it uses other representations to manipulate the animated world (e.g., tables, algebraic notation). `SimCalc MathWorlds`® was evaluated [61] with large groups of teachers and students of different backgrounds and skills (p. 836). Interestingly, `SimCalc MathWorlds`® was specially designed to integrate technology, curriculum

[3] `http://www.kaputcenter.umassd.edu/projects/` and `http://math.sri.com/`

and teacher training. Some of the hallmarks were that students using the software were able to make predictions, students were able to make sense of *"conceptually rich mathematics in their experience of familiar motions"* (p. 839), and it facilitated students' learning of complex concepts as well [62].

Thus, dynalinking can happen in different forms: the user could interact with the object/situation and observe how the abstraction is affected; or the user could interact with the abstraction and observe how the object/situation is affected. Our work considers this property for our learning environment.

Once a learning environment has been set up, how are we going to assess students' knowledge of line graphs? That is, how do we know that, after the intervention, students' interpretation is (close to) what we expect them to understand? In the next section, we review in more detail the current methods to assess students' line graph interpretations.

2.6 Assessment of students' diagrammatic knowledge

As we mentioned before, we need to be able to assess students' knowledge of line graphs in order to measure the effectiveness of our learning environment. In this section we will start by stating what we understand by graph comprehension and we continue with an overview of some of the common assessment methods of students' graphical knowledge. Afterwards, we point out some of the limitations that those methods convey, and this will take us to our second research question.

2.6.1 Students' level of interpretation

We start by borrowing the definition of graph comprehension given by Friel, Curcio and Bright [32]:

> By *graph comprehension*, we mean graph readers' abilities to derive meaning from graphs created by others or by themselves (p. 132).

The focus of graph comprehension in the literature has been mostly on *"questions that graphs can be used to answer"* ([32], p. 130). Friel et al. [32] provided a summary of some of the levels of questioning (that seems to be in general agreement with other research) and which Curcio and Smith-Burke [26] called "reading the data", "reading between the data" and "reading beyond the data". In Wainer's [72] view, those three levels of comprehension can be explained as: an elementary level of questions for data extraction, an intermediate level for observing trends, and an overall level for understanding the deep structure of the data (p. 16).

Examples of the questions that correspond to those levels for Figure 2.1b are:

- At what time the sound became louder? *(Reading the data)*
- Between 17:30 and 18:00, how did the sound change? *(Reading between the data or interpretation)*
- Which is the pattern of sound or how is it predicted to change? *(Reading beyond the data)*

In this thesis, we focus on reading between the data or interpretation. In particular, for *interpretation* we will understand:

> a progressive integration of the various pieces of information conveyed by the graph with the underlying situational background ([38], p. 3.6).

2.6.2 Methods used to assess interpretation of diagrams

The literature shows that three methods to assess students' line graph interpretations are typically used: multiple-choice (e.g., [38], [37], [18], [44], [11], [53]), free-response (e.g., [10], [42]) and graph production (e.g., [10], [18], [11]) instruments.

Multiple-choice instruments (MCI) might be one of the most common methods that is still widely used in research (e.g., [44], [26]). Although this type of instrument helps us to observe specific pieces of students' knowledge, it also brings some limitations. For instance, target-items are usually limited in number and offer very specific views/interpretations of the graph. Therefore, this method disregards other possible varieties of students' conceptions besides the ones presented there. Consequently, it provides us with limited insight into students' depth of knowledge. In other words, through multiple choice instruments we might get an insight into what specific knowledge structure was cued more strongly, but an MCI is limited in the sense that we do not know which other knowledge structures/p-prims are linked or explain the selected answer.

Free-response instruments could provide a better insight into students' knowledge since students could express their actual conceptions. Thus, this method could give us a richer view of their knowledge. However, assessment could prove more difficult to do than with multiple-choice instruments. For example, a student selecting a correct answer can be identified by the selection of the target item in an MCI, whereas in a free-response instrument, the examiner has to read, interpret and judge whether a student's response has offered enough evidence of understanding and interpreting the graph correctly – a process which could be time consuming and/or subjective. In addition, free-response instruments could be challenging for young students who might still not be very confident at writing.

Graph production could offer an alternative to free-response examinations. Nonetheless, as it was noted before (section 2.3.1) students might need some previous experience when constructing different types of graphs and diagrams, particularly very young students.

In this thesis we used the multiple-choice instrument, but we also attempted to develop an alternative instrument to observe young students' graphical knowledge. In regard to this, research by Cox and Grawemeyer [24] is relevant. They used a card-sort task to observe their participants' mental organisation of external representations. Their participants were asked to categorise a group of cards and to name their categorisations. Cox and Grawemeyer observed that subjects who performed better in problem solving and reasoning tasks, organised their cards according to their semantics rather than superficial similarities ([24], p. 95). This work is very close to our aims, since we wanted to observe students' graphical knowledge and whether WYSIWYG (section 2.4.2) was cued across a variety of items.

Thus, inspired by Cox and Grawemeyer's [24] research and we used a graphical decision task designed to get a better insight into students' depth of knowledge (more details are in Section 3.1.2).

Works related to the assessment of the "graph-as-picture misconception"

Most of the research we reviewed assess students' pictorial interpretations using multiple-choice instruments, and free-response examinations or graph production. However, the most frequent method used seems to be multiple-choice questionnaires. In particular, Janvier's task has often been used to elicit pictorial interpretations (e.g., [70], [45]). Thus, in this section we give a brief summary of Janvier's research.

Janvier [38, 39] was one of the first researchers to look into more detail on students' graph-as-picture interpretation (GAPI). He was actually concerned with the role of situations in mathematics and conducted several in-depth interviews with secondary school students to observe whether the use of situations guarantees *"'concretisation' of abstract notions"* ([39], p. 113).

Janvier's aim when developing his task was to:

> analyse the variety of mental processes involved in the task in regard to the global features which the graph included, and the role played by the familiarity of the subject with the underlying situation ([38], p. A3.4).

His assessment task was used as an assessment of students' ability to derive meaning from the graph and to assess whether students are deceived by the graph with a picture of the situation. Janvier presented students with Figure 2.6 and instructed them:

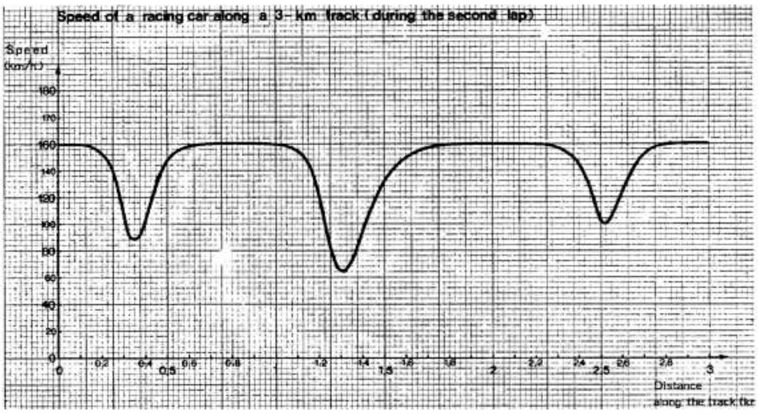


Figure 2.6: Janvier's graph interpretation task. Source: [38] (p. A3.4).

> As the heading shows, this graph tells how the speed of a racing car varies along a 3km track during its second lap. The speed is given in km/h. This axis (pointed along with the forefinger) gives you the distance along the track from the starting line. So, this graph gives you the speed of the car anywhere along the track. It gives you, for instance, the speed of the car after 1km, after 1.8km and 2.4km ([38], p. A3.5).

Afterwards, the student was asked *"Can you tell me, from the graph, how many bends there are along the track on which the car was driven?"* ([38], p. A3.5). Depending on the students' answer, the examiner continued asking for their interpretation or provided hints to the students to help them solve the problem.

Janvier's analyses were mainly qualitative since the procedure for administering his task was performed using in-depth interviews. One of the issues Janvier noted was that students' past experiences could have influenced how students used the graph. For example, his male participants seemed to have had some experience with racing car games and Janvier' suggested that this could partly explain why boys were better performers than girls in his task. This suggests that future evaluations could be made using gender-neutral contexts.

In terms of the role of the situation, Janvier noted that, sometimes, the concrete features of the situation could be distractive and could make the use of the abstraction more difficult ([38], p. 9.44). However, he believed that:

> ...down-to-earth handling of the idea of speed carried out with the help of a graph is most valuable for the pupils and that it enables the perceptual distractor to gradually "brush-off" as part of the concept evolved ([38], p. 9.5).

That is, he suggested that using tasks *"involving sensory-motor responses"* ([38], p. 3.13) could support understanding of the abstraction, but this is an issue we investigate with our first research question.

2.6.3 Accuracy describing students' diagrammatic knowledge state

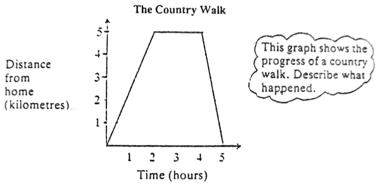
The use of the term "graph-as-picture misconception" seems to be used indiscriminately for a group of different students' interpretations. For example, Bell et al. [10] presented examples of some students' interpretations of a distance-time graph showing a "country walk" (see Figure 2.7a). One of the typical pictorial answers is interpreting the graph as a picture of a hill. Nevertheless, other types of responses were also given by pupils. For instance, one of the students interpreted it as follows:

> What happened was he or she walk NE for some distance, then it was one straight road going E then they changed direction straight away to go slightly S ([10], p. 56).

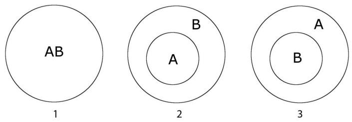
This answer shows the student interpreting the graph as if it were another representation, as if it were a map instead of a "pure" picture. More examples of this type of interpretation can be found in [42]. Another student gave the following interpretation for Figure 2.7a:

> He started from 0 and then went [*sic*] a walk and walked round in a square and ended up at number 5 ([10], p. 56).

This time, the graph is not interpreted as a picture of a real world object nor it is fully interpreted as a map (as in the previous case). However, this case has been classified by Bell et al. [10] as a "graph-as-picture misconception" too.



(a) Line graph: the country walk.



(b) Euler's circles: all As are Bs and some As are not Bs.

Figure 2.7: Interpretation tasks using a line graph and Euler's circles. Sources: Line graph from [10] (p. 56) and Euler's circles from [20] (p. 437).

Although a GAPI has usually been reported when using line graphs, it is possible that it happens when interpreting other external representations as well. Cox [20], for example, presented a case where students gave what he termed "island responses" when interpreting Euler's circles. Cox explained that some students could interpret two of the Euler's circles in Figure 2.7b (number 1 and 2) correctly as "all As are Bs"; however, they could subsequently interpret Figure 2.7b, number 2, as "some As are not Bs" – as if the circle A represents an *"'island' of As in a 'sea' of Bs"* ([20], p. 438).

Overall, it seems that the term "graph-as-picture misconception" has been used for a variety of students' answers. It has been used to describe students' "map" and other vague interpretations. This highlights that students' knowledge has not been assessed appropriately or accurately, but a knowledge-in-pieces perspective (see section 2.4.2) could help us to find a better way to assess students' knowledge.

2.7 Research questions compiled

The thesis presents work around two questions. The first question relates to whether active or passive learning environments could help improve students' knowledge of line graphs (Section 2.5.3). The second question is related to how to assess students' knowledge of graphs (Section 2.6.2 and Section 2.6.3). Both questions are tightly related since we need to be able to track how the learner's knowledge changes when a learning activity occurs.

One way to look at the relationship between those two questions could be explained using Friederick Reif's perspective of teaching and learning ([59], Figure 2.8). Fitting his model with this research, we could assume that students come with an initial state S_i and we want them to get into an ideal final state S_f. The transition from S_i to S_f is affected by some learning intervention (L). The diagram in Figure 2.8 exemplifies these three aspects. Focusing our research on line graphs we find that:

- S_f represents the students' state after L has occurred. For example, S_f could be

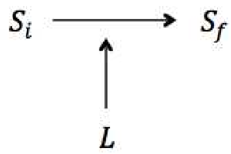


Figure 2.8: Reif's model adapted to this research ([59], p. 4).

the state where the student is able to read axes or operate values of line graphs. However, this research aims that, after the application of L, students should be able to interpret line graphs appropriately. We focus on this aspect of students' comprehension since students have more difficulties achieving it (other pieces of research show that students tend to be more successful when reading axes or operating values of the graph).

- S_i represents a previous state of S_f; that is, the students' knowledge of line graphs before the application of L. If in state S_i students do not interpret line graphs appropriately then S_i represents students' deficient performance. This research focuses on one of the more predominant difficulties when interpreting line graphs, which is known as the "graph-as-picture misconception". Although students could have other difficulties when interpreting graphs (such as the slope/height confusion, relating two variables, reading the graph point-wise, etc.), we focus on students pictorial interpretation because: (1) it is one of the well-known problems when interpreting line graphs; (2) research suggests that a number of pictorial-like interpretations may happen and we may need a more effective way to assess students' knowledge; and (3) there is current literature on this subject that provides a base to this research.

- L represents the scaffold that should help students to get from S_i to an ideal state S_f. In this research, L represents the learning environment(s), which we test.

In summary, our aim is to help students to improve their interpretations of line graphs (particularly students who give a type of graph-as-picture interpretation); therefore, we need to find an effective learning environment (L) that could help students to improve their understanding of line graphs and we need an effective method to assess students' pictorial interpretations in state S_i and in state S_f.

Thus, considering the specific areas of this research and our interest for students giving GAPIs, our research questions are as follows:

(1) Will an active learning environment be more effective to improve students' interpretations of line graphs than a passive learning environment?

(2) How can we assess students' interpretations of line graphs?

We will start to answer these questions by developing some of the core tasks of our experiments, which are presented in the next chapter.

Chapter 3

Pilot Study and Technology Development

This chapter presents a pilot study and reviews the interactive learning technologies that we considered using in the subsequent experiments. The pilot study demonstrates some of the tasks used to answer our two research questions: (1) Will an active environment be more effective to improve students' line graph interpretations than a passive learning environment? (2) How can we assess students' interpretations of line graphs? In particular, we present details of the learning environment's design (which was an improvement of Janvier's Racing Car activity [38] and which implemented *dynalinking*). We also describe an alternative method to assess students' graphical knowledge (a card-sort task, which we wanted to use for observing students' interpretations in a rich way). Afterwards, we present our observations regarding the experiences of three students who took part in this pilot study. In order to provide an embodied learning experience to students, a large interactive touch-screen display was used as an interface. Additionally, following [22], the learning system was designed to function as a "technologically-enhanced research" tool. That is, pupils' interactions with the system were richly logged, screen recordings were stored and students were also videotaped (with their guardians' consent). These sources of data were triangulated in order to identify students' difficulties and to evaluate our tasks' design.

The second part of the chapter describes the different interactive learning technologies we considered. It also mentions to what degree those technologies met our requirements and how well they facilitated the interaction with our tasks. Outcomes of this summative usability study not only helped us to pick the most adequate interactive learning technology for our activities, but they could also inform other researchers to select an appropriate interactive system for their tasks.

At the end of the chapter, we provide a summary of the lessons we learned from the pilot study and from our review of the different interactive technologies.

3.1 Pilot study

In this section we describe the design of a Racing Car activity, a card-sort task and a graph comprehension questionnaire. We gave these three activities to three participants and evaluated our tasks' design based on the participants' behaviours.

3.1.1 Aim of the study

The tasks presented here are adapted from Janvier's [38] paper-and-pencil tasks, Curcio and Smith-Burke's [26] graph comprehension questionnaire, and Cox and Grawemeyer's [24] card-sort task. Although Curcio and Smith-Burke's questionnaire was administered to young students, Janvier's and Cox and Grawemeyer's tasks were originally designed for secondary and tertiary school students.

Thus, our aim was to improve these tasks to be used with primary school students and within the context of line graphs. In addition, we also wanted to administer the tasks through an interactive computerised system. The evaluation should provide indicators towards the applicability of these tasks to answer our research questions.

3.1.2 Development of the tasks

A description of each task's characteristics is presented below. Note that these tasks were specifically selected to help us address each of our research questions and to help us observe which students give pictorial-like answers.

Learning environment

In order to answer our first research question, we needed to create a learning environment that could be used either actively or passively (see Section 2.5.2). The literature shows that Janvier's [38] paper-and-pencil tasks have been frequently used to observe and improve students' interpretations of line graphs (e.g., [45]). Thus, we implemented an interactive version of Janvier's Racing Car activity [39] that could be used actively or passively. If we want to create an effective learning environment that could help students to improve their knowledge of graphs, which characteristics should we consider for it? The literature on this subject is discussed in Section 2.5.4 and Section 2.5.5, but some of the main studies to consider are by Vogel, Girwidz and Engel [70], Rogers, Scaife, Aldrich and Price [60], and Janvier [38, 39] (see Section 2.5.4 and Section 2.5.5).

Vogel et al. [70] and Rogers et al. [60] designed multimedia-learning environments to help students connect an abstract-metaphorical representation to a concrete representation. Their environments highlighted the use of *supplantation* or *dynalinking*. That is, they made explicit how the features of the representation co-vary with the features of the situation when the situation (or the representation) is modified. Thus, considering those researchers' work and using Scaife and Rogers' [63] External Cognition framework, we incorporated *dynalinking*.

The Racing Car task In our Racing Car activity, students used an interactive touch-screen and moved an image of a "car" around a track. A corresponding speed-distance graph was concurrently plotted alongside the track to make the correlation between the car movement and the plotted line more salient. Dynalinking took the form of (1) colouring the track behind the car as it progressed around the track and using corresponding colours on the graph axis; and (2) adding a "speedometer" on the top of the car. It was expected that dynalinking would highlight visual features of the graph and act as a cognitive aid [46, 60]. A screenshot of the software is in Figure 3.1. Other characteristics where considered for the design of the Racing Car activity too, and are explained next.

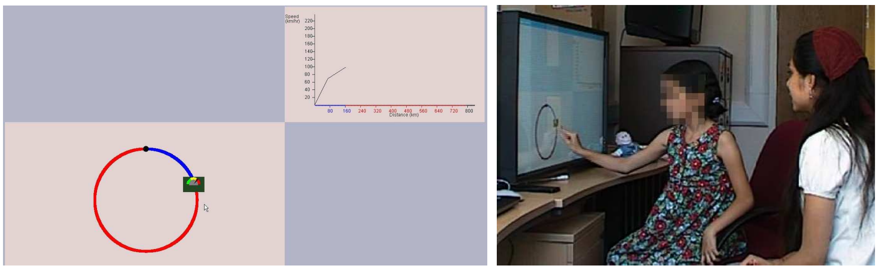


Figure 3.1: On the left side, a screenshot of the Racing Car activity: the track is on the bottom-left area and the graph is on the upper-right area. The "car" is the green rectangle over the track (the "speedometer" is on the top of the "car"). The image on the right side shows a student doing the activity.

Layout Note that the track and graph are positioned diagonally. This decision was taken after considering Avons, Beveridge, Hickman and Hitch's [5] research. They found that positioning the track along the graph's axes could affect the way the graph is interpreted (see Section 2.5.5 for more details). Thus, it was decided not to put the track next to the sides of the graph, but along its diagonal in order to avoid its position to be a source of confusion for students.

Speedometer The speedometer featured only 3 speeds to facilitate its use: slow (green), medium speed (yellow) and fast (red). We did not want the speedometer far away from the car or as a representation in another part of the screen that the student had to consider because that would require him/her to simultaneously attend to more parts of the screen. We did not want it "floating" on the side of the car too since it would occlude the track at some points (and this means that the speedometer would need to change position during the race). Thus, we put the speedometer on the top of the car. We also made the "car area" large enough so that the student could move the car and see the speedometer at the same time. A close-up of the speedometer is in Figure 3.2 (the car is represented by the dark-green rectangle).

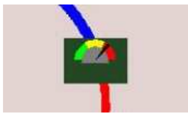


Figure 3.2: A close-up of the speedometer used in the Racing Car pilot activity. See Figure 3.1.

Sound The car's volume increased if it was moved fast and its volume decreased if the car was raced slowly. We added this as another form to support the students' making sense of the graph (e.g., the student could attend to the graph and hear the car's speed simultaneously). According to Mayer's Cognitive Theory of Multimedia Learning, this might not increase students' cognitive workload since the visual and auditory information would get processed by different channels [47].

Tracks The activity tracks (AT) and trial tracks (TT) we used are in Table 3.1. The tracks used in Janvier's [38] original tasks were similar to AT2, TT2 and TT3 and we included other "closed" and "open" tracks to the activity too. For the trial tracks we wanted students to practise moving the car (1) on a track with no bends, (2) on a track with no straight parts, and (3) on a track with sharp/wide bends. The activity tracks showed (1) a straight line (to show students the difference between the generated plotted line and the shape of the track); (2) a closed track with two wide bends (to allow students to see what the plotted line looks like when travelling on straight and curved parts of the track); and (3) a graph with wide and sharp bends (to allow students to observe how the plotted line changed when travelling along wide/sharp bends of the track).

Track sequence:	#1	#2	#3
Trial tracks (TT)			
Activity tracks (AT)			

Table 3.1: Activity tracks and trial tracks for the Racing Car pilot activity.

Replay In order to assure that students could observe how their experiments changed the graph, a replay of their "races" was done after each trial. This design allowed using concurrent verbal ("think aloud") [30] and retrospective debriefing (CV-RD)[66] protocols. Thus, it permitted us to observe students' thoughts and strategies at various stages.

Touch-screen technology As mentioned in another study [34], it was important that students could use their natural movements and *"receive kinaesthetic and proprioceptive feedback from their actions... without mediating artifacts [sic] such as computer mice, keyboards interposed between the learner and the material being learned"* (p. 43). Beichner [9] suggested that this type of interaction creates an event-graph unit, which it is easy to manage in working-memory and to keep in long-term memory as a single entity (p. 804–805). Mokros and Tinker [53] supported the use of a "*...real-time link between a concrete experience and the symbolic representation of that experience"* (p. 381). Janvier also indicated it would be valuable for pupils if his tasks could use sensory-motor responses ([38], p. 9.5). We wanted to keep this relationship between the physical action and the abstract representation so that

> motor memories of actions (fine and gross motor movements of the arm, hand and fingers) and kinaesthetic feedback from friction of the fingertip against the screen become associated in the student's mind with changes in speed of the racing car being "driven" by the student ([34], p. 43).

Assessment of graphical knowledge

As mentioned in the previous chapter, there is evidence that primary school students are able to understand line graphs, but how can we assess their line graph interpretation?

Studies that assess pupils' graphical knowledge tend to focus on students above primary school level (e.g., [71]) and they frequently use methods such as multiple choice questionnaires and free-response instruments to observe students' graphicacy (see Section 2.6.2). However, considering that knowledge structures are complex (see Section 2.2), employing multiple-choice instruments seems to give us with partial insight into the students' interpretation; free-response instruments could provide greater insight into the students' depth of knowledge, but they could be difficult and time consuming to evaluate. Another alternative method to obtain insight into some of the students' mental organisation (and which evaluation could partly be automatised) is through a card-sort task similar to the one done by Cox and Grawemeyer [24].

Cox and Grawemeyer [24] used a card-sort task to find out about people's organisation of graphical knowledge. They asked participants to categorise a group of cards (that showed different types of external representations, ERs) and to name their categorisations afterwards. They also asked participants to identify the function of various forms of external representations. This technique allowed them to assess their subjects' depth of knowledge at the level of recognition, naming and semantic knowledge. They found that *"expert ability to use ERs in problem solving and reasoning is associated with more accurate naming of ERs and a tendency to create categories on the basis of ER semantic distinctions"* ([24], p. 95). In other words, non-experts' categorisations seemed to be influenced by visual factors of the cards and they seemed to *"respond more to superficial aspects of ERs, particularly to spurious visual features"* ([24], p. 95).

In this study, Cox and Grawemeyer's [24] approach was modified in order to be used with young students. A card-sort task was created and designed to assess students' graphical knowledge and to observe whether pupils failed to recognise a representation as an abstraction, whether students could remember the name of the representation or whether they could understand what the representation meant. More details of this task are presented next.

Cards and categories It was possible that students could successfully recognise well-formed representations, but they could fail to do so if the representation showed distractive perceptual features (e.g., "decorations" on the graph) or if the representation had flawed characteristics (e.g., missing labels). That is, comprehension of the graph might be obstructed if important parts of the representation were overlooked/disregarded as a consequence of, for example, pictorial features. Thus, the items in the card-sort task contained examples of (1) diagrams, (2) diagrams with missing features, and (3) pictures or pictures looking like diagrams (see Figure 3.3). That meant that instead of allowing participants to create their own categories (as Cox and Grawemeyer [24] did), we specified the categories in advance. Therefore, during the task, three category areas, to place the cards, were shown on the screen.

Students were required to assign each of the cards presented on the screen to one of the three areas. We did not want too many cards to be sorted per trial; thus, we decided to show only three items at one time (see Figure 3.4). We kept a reduced number of cards (and categories) because for each additional card the number of comparisons the student had to make increased[1] and, therefore, the complexity of the task increased (this also could augment the time to finish the trial and could make the analysis more difficult to carry out).

We kept the cards large enough so that students could clearly see their details. It was not possible to overlap the cards so that pupils could see all three at all times. It was not possible to place two cards on the same area (per trial) as well. Note that the representations shown on the cards were drawn by hand and some of them where selected from school textbooks or UK National Curriculum documentation.

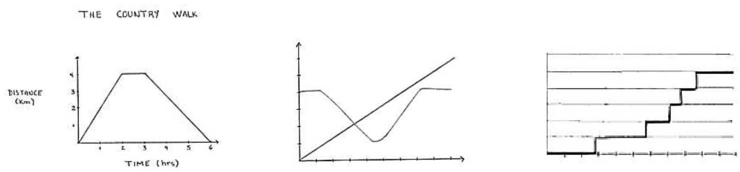


Figure 3.3: Examples of the items used in the card-sort task. From left to right: a diagram, a diagram with missing labels, and a picture that looks like a diagram.

Overall, we expected that by observing students' categorisations (either successful or not), we would be able to see the students' degree of knowledge of the representation shown[2].

Wording for the instructions In order to keep the task accessible to all participants in our target range (3rd to 6th school year) it was necessary to use terms that all of them were likely to be familiar with. Hence, words that do not relate with scientific topics where chosen for the instructions (e.g., the word "graph" might be considered to be related to numbers and it might bias students' decisions); that is, it was decided to

[1]The number of ways to position the cards was $n!$, with $n = number\ of\ cards$. For example, 2 cards (A and B) can be sorted as AB or BA ($n! = 2! = 2 * 1 = 2$); 3 cards (A, B, C) could be placed in 6 different ways ($n! = 3! = 3 * 2 * 1 = 6$): $ABC, ACB, BAC, BCA, CAB, CBA$; etc.

[2]There were six representational forms used in this task, but since this work is concerned with students' understanding of line graphs, we only focus on this representational form.

avoid using words like "diagrams", "graphs" and "charts". Adjectives and nouns referring to the *content* of the cards were preferred rather than those describing the *appearance* of the card. Thus, we tried to use words related to *unusual*, *inappropriate* or *non-meaningful* characteristics of the cards' contents, but using a more child-friendly vocabulary. The words used here (*crazy*, *ridiculous*, *nonsense*) were chosen because very young children are familiar with them (examples of this can be found in songs and poetry for children) and they reflect current youth idiomatic language. Although, those words can be interpreted subjectively, we used them in order to observe if this kind of vocabulary could be used when working with young students (i.e., can we use instructions without using words like "graph" and without an explicit mention of the categorisation in order to avoid biasing students' decisions?).

We also asked participants to order the cards according to how *not*, *a little*, or *most* unusual things were on the cards in order to see how they would organise each card with respect to the others. For example, if we were showing a well-formed graph, a graph without labels, and a picture, then the way students organise the cards could tell us more about their graphical knowledge. For instance, if a student sees the graph-without-labels as more unusual than the well-formed graph item then this could show that the student is identifying some features that are not what he/she expects; or if he/she classifies the graph as more unusual than the picture, then this could show that the student might not recognise the representational form; etc.

Thus, we related our "not unusual things" area to the cards that show well-formed graphs, the "a little unusual" area to the graphs with missing features, and the "most unusual" area to the pictures that look like graphs. Additionally, in order to evaluate the students' performance, the students' classifications were compared to our "expected" answers as well (see Appendix A).

Technology-enhanced research Students' answers and sorting time were logged, utterances produced were recorded and screen and video recordings were taken (with their guardian's consent). The task was administered using a touch-screen device too.

Overall, it was expected that this approach would facilitate the student's experience, would allow collection of a wide range of data (e.g., decision latencies, screen recordings, utterances) and it would allow us to observe other students' behaviours. For example, cards might have been placed onto the target areas in order to consider what certain classifications would look like in that order, but if that placement prompted a conflict then the cards would be replaced somewhere else.

Line graph interpretation

Since we were interested in supporting students whose line graph interpretations could easily activate the WYSIWYG p-prim (see Section 2.4.2), we needed to select a group of students who interpreted a line graph pictorially. In general, evidence of students' pictorial answers have come from one of the following three forms of tests: multiple-choice instruments (e.g., [38], [37], [18], [44], [11], [53]), free-response instruments (e.g., [10], [42]) or graph production (e.g., [10], [18], [11]), but one of the most successful methods to trigger the WYSIWYG p-prim is through multiple-choice instruments (e.g., [11], [42]). Note that this method does not assure us that we will get a group of *all* the students who interpret the graph pictorially, but it could be an effective method to obtain our sample group.

A modified version of Curcio and Smith-Burke's questionnaire [26] was used to assess line graph comprehension of primary school pupils. The activity presented a line graph describing a "journey" (see Figure 3.5) and involved four multiple-choice questions that assessed different facets of a student's comprehension. A "journey" topic was chosen as it was thought not to be gender biased. Comprehension questions required the student to read the axes (x and y), perform operations (e.g., *"How much did Peter walk between 3 PM and 5 PM?"*) and interpret a graph (see Section 2.6.1). The interpretation question contained items that make reference to the graph as a picture of the situation (e.g., *"Peter walked up a hill and then went down the other side."*).

The questionnaire is in Appendix A, but question 4 and its target items are below (the correct answer to the question is in bold).

4. Which of the following sentences describe what happened in Peter's trip?

   Peter walked more in the last 3 hours than in the first 3.

   Peter walked faster and then went down a hill.

   Peter walked up a hill and then went down the other side.

   Peter walked faster in the last 3 hours than in the first 3.

Note that the target items for question 4 included the concepts of speed (e.g., "walk faster"), time (e.g., "last 3 hours"), distance (e.g., "walked more") and the "pictorial" interpretation of the graph (e.g., "went down a hill"). Thus, looking at the target items for question 4 from top to bottom, the concepts presented are distance-time, speed-pictorial, pictorial-pictorial, and speed-time. The intention was to make the pictorial-pictorial option more salient from the other options and, at the same time, to capture instances of an alternative pictorial interpretation of the graph.

3.1.3 Description of participants

Three young students (male, 8 years old; female, 9; and another male, 11) participated in the activities described above. They were in the 3rd, 4th, and 6th primary school years, respectively. They will be identified as S8, S9 and S11.

This pilot study was conducted in a quiet laboratory. Screen recordings were made of all sessions using Techsmith Camtasia Studio 4. Sound recordings were made only with S8 and S9. Video recording was done only with S9 (with informed parental consent).

3.1.4 Procedure

Students performed the activities through a touch-screen 40" NEC Touch Screen MultiSync LCD4020. All participants did some drawings on the touch-screen before starting the tasks in order to help them be familiarised with it.

In all the activities instructions were provided on-screen and they were either read by the participant or by the experimenter. No feedback was provided for any of the tasks. During the card-sort task and the Racing Car activity, students were encouraged to talk as much as they could, but for the graph comprehension questionnaire there was no intervention from the experimenter in order to avoid interfering with students' answers. In addition, students were allowed unlimited time to complete the tasks.

The three activities were administrated in the following order: (1) the card-sort task, (2) the graph comprehension questionnaire, and (3) and the Racing Car activity.

Card-sort task

Instructions were presented at the start of the activity and they were as follows:

> Look at the cards CAREFULLY and choose: which has THE MOST nonsense, crazy or ridiculous things on it, which has ONLY A LITTLE nonsense, crazy or ridiculous things on it, and which DOESN'T have any! Try to speak aloud whatever that comes into your mind.

For every trial, three cards were selected randomly from the data corpus and were placed in random position within the screen. Students were required to place the cards onto one of three different screen areas. The areas were placed on the right side of the screen and they contained, from top to bottom, the text *"MOST nonsense or ridiculous things"*, *"ONLY LITTLE nonsense or ridiculous things"*, and *"DOESN'T HAVE nonsense or ridiculous things"* (see Figure 3.4). Once the student finished sorting the cards, the button “Done” (at the bottom of the screen) was pressed; this showed the next trial to the student.

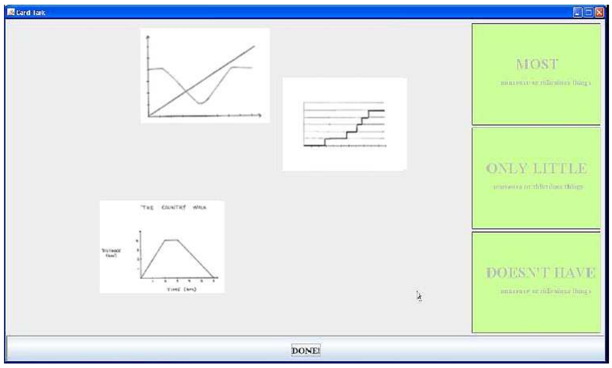


Figure 3.4: Screenshot of the card-sort task.

Graph comprehension questionnaire

The questionnaire consisted of four questions. Each question was presented on a different screen. Each screen showed the line graph in Figure 3.5 on the left hand side and the question in turn on the right-hand side. The first three questions asked students to read the horizontal axis, the vertical axis and to perform an operation on the line graph, respectively. After that, there was a forced pause indicating: *“Please think about the graph and then continue”*. Finally, the interpretation question was presented.

Additionally, before starting the activity, students were presented with the following context:

> Peter is a kid that lives in a very very flat town. He likes to walk and he decided to go for a walk to town. The next screens show a graph about Peter’s trip. Read the questions and answer them using the graph. You can take as much time as you need.

Note that the idea of a “flat town” was emphasised by repeating the word “very” twice.

Racing Car activity

At the start of the activity, children were told what the speedometer did and how the colour in the path was matched in the graph. They were instructed:

> With your finger move the car along all the track. At the same time try to see how the graph changes when you speed up or slow down your movements. And speak aloud everything that comes into your mind.

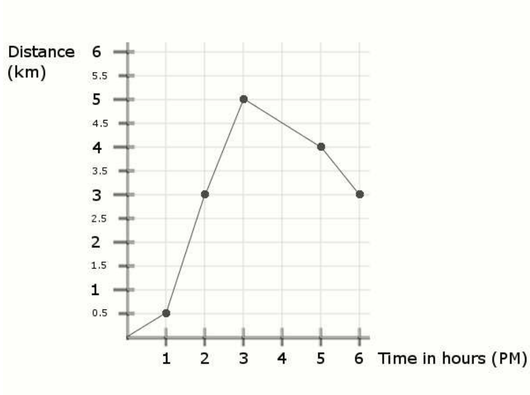


Figure 3.5: Graph line used for the graph comprehension questionnaire.

Afterwards, students practised with three different tracks (see Table 3.1, first row) of the Racing Car activity before starting the proper task, which showed three other tracks (see second row of Table 3.1 and Table 3.2).

3.1.5 Results: students' experiences

Details of students' experiences are presented here. The order we present our observations is according to our research questions rather than the order we showed the activities to students.

Learning environment: Racing Car activity

Table 3.2 shows the graphs produced by the participants for the trial phase and for the activity. Students' strategies during the activity are discussed next.

S8's strategies For S8, the activity resembled a game. S8 showed high enthusiasm either when racing the car or when watching the replay of his race. For instance, during the replay he shouted: *"do it faster, faster, faster, faster..."*. That is, in spite of the car being represented as a green rectangle, he did not have difficulties looking at that representation of the track as a "real" situation. For example, while doing the activity he uttered: *"Let's go extra fast! Come on then! Make it work this piece of tin..."*.

During the trial tracks #2 and #3, and in all the activity tracks, he expressed his plans for *"going red"* or *"going speedy"* – that is, his utterances signalled that S8 was using the dynalinking features of the software as well.

With respect to S8's interpretation of the graph, positive indications of appropriate interpretations were observed. For instance, after racing the AT3 (Table 3.2), the experimenter asked (pointing towards the negative slope of the graph generated): *"What happened here?"* S8 answered:

> I went slow... and that's when I went really slow. So when it goes higher, goes higher speed...

Additionally, audio evidence showed that S8 was engaged in reading the graph. For example, S8 commented:

> It's quite hard... you... you have to go down and... six hundred...

However, there were a couple of issues with S8's experience. First it seemed that there was a tendency to think about the track as a picture of an object when S8 identified the last track as "a bell": *"come on bell!"* Second, it was noted that on some tracks, S8 had problems racing the car how he would have liked since the interface did not seem to respond as fast as his movements. However, when he managed to drive the car skilfully, he produced an "ideal" plotted line: see the graphs for AT1 and AT3 (they show that the car went fast in straight segments and slow around corners).

S9 and S11's strategies In the case of S9, she seemed to view the activity as a game too. When S9 was asked *"what are you trying to do?"* her answer was *"I'm trying to make it a bit faster"*. Her utterances showed that she was looking at the speedometer; that is, she was using the dynalinking feature too:

> **Experimenter**: How did you know that it was fast?
>
> **S9**: Because the red needle...

S9's comments showed that she was able to interpret the graph as well. Her experience during TT3 was particularly interesting: It allowed S9 to (perhaps) relate the concept of "stopping" with "the bends of the track":

> **Experimenter**: Why was it more difficult?
>
> **S9**: Because it stopped when it went to turn...

S9 had similar problems as S8. Her comments reflect that she was relating some tracks with pictures of objects. For instance, when looking at the last activity track, S9 said: *"this is a shoe shape"*. Additionally, S9 had some problems using the touch-screen when doing the activity – this can be noted when looking at the plotted line produced when she raced the track AT2 (see Table 3.2).

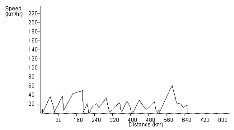
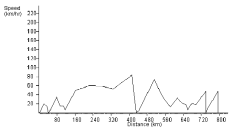
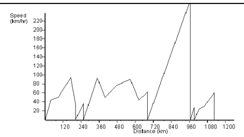
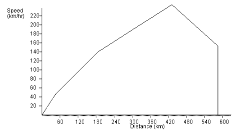
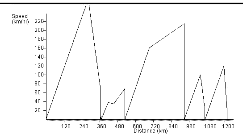
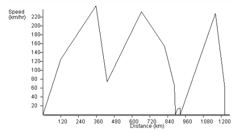
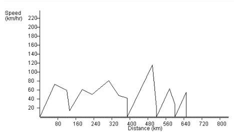
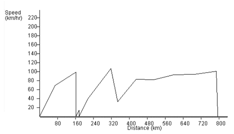
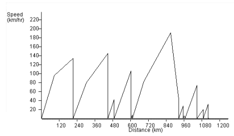
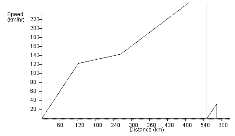
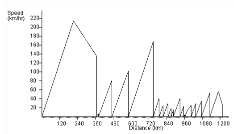
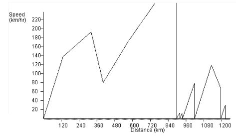
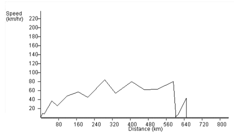
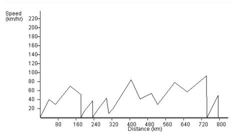
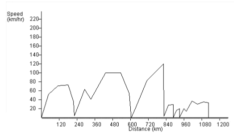
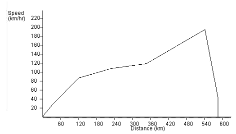
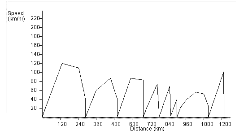
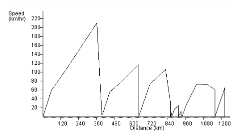
	Trial tracks (TT)			Activity tracks (AT)		
						
	#1	#2	#3	#1	#2	#3
S8						
S9						
S11						

Table 3.2: Graphs generated during the Racing Car activity. The tracks were presented to the students from left to right.

S11's utterances were not recorded, but notes were taken during the activity. During his races, S11 watched the graph closely, and he was able to correctly interpret it when he was asked. For him, the interface needed more work and the car needed to look like a real car. In particular, he observed a delay between the movement of the car and the plotted line of the graph, which indicated that the response of the interface needed to be improved.

Assessing graphical knowledge: Card-sort task

Table 3.3 shows how the participants sorted the line graph items during the task. Note that the second column for S8 is missing because S8 quit the activity after finishing the first round (he seemed to find the task somehow challenging). The order to sort the cards that was considered more appropriate is shown in Appendix A.

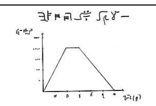
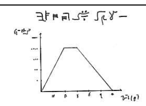
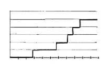
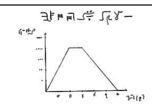
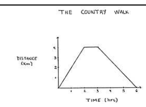
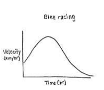
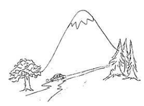
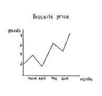
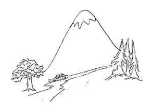
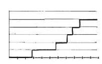
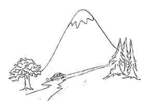
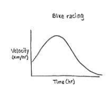
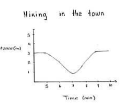
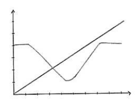
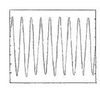
		S8	S9		S11	
Type	Order	Round 1	Round 1	Round 2	Round 1	Round 2
Line graphs	Most					
	A little					
	Not					

Table 3.3: Students' answers for the card-sort task.

S11's classifications For the first round, it seemed as if S11 judged the cards partly by their *content*. For instance, S11 placed a full-labelled line graph in the "most nonsense" area: perhaps he noticed the card showed a well-formed graph, but the content looked suspicious so then S11 decided to put it in the top target area.

However, for the second round S11 might have sorted the cards by *familiarity*. For instance, one of the card items was placed as "not nonsense" since it looked like "the screens you see in hospitals" (e.g., a heart rate hospital monitor).

In both cases, S11 put the full-labelled graphs on the top, but for different reasons. Since his classifications were not consistent in both rounds, it is difficult to assess his knowledge of line graphs with only these items.

S9 and S8's classifications S9's classifications were more consistent than S11's classifications. S9 positioned most of the cards with all the labels at the bottom. That is, she seemed to consider the labels of the graph when categorising them and was able to recognise when the graph seemed well-formed. However, unknown features on the graph could have been a source of distraction for her (e.g., a well-formed graph that shows labels with unknown characters was categorised as "most unusual").

However, for S8, the task was more difficult. Different factors could have made the task more challenging for S8 such as: clarity of instructions and use of the interface (he found it difficult to move the items easily on the screen). All these aspects need to be improved if this task is going to be addressed to young students.

General comments on the card-sort task Overall, it seems that the card-sort task needs to be friendlier for young students and it should be modified to facilitate analysis. That is, the participants' answers appeared to provide some idea of the strategies they used. However, it is not clear whether it is possible to obtain a clearer insight of students' knowledge with this layout of the activity.

We could improve the task by increasing the number of cards students see in order to obtain more utterances from pupils. We could reduce the number of categories to facilitate the analysis too. Instructions could be improved by using the actual name of the categories we are interested in – for example, students were familiar with the term "graph" when they were asked at the end of the session. Employing specific vocabulary could help to reduce ambiguity when interpreting students' answers.

Graph comprehension questionnaire

Table 3.4 shows students' correct or incorrect answers to the questions presented. Overall, the questionnaire proved useful to observe some of the students' pictorial interpretations. S8 and S9 responded with one of the pictorial target items; they chose the option *"Peter walked up a hill and then went down the other side"*. However, it was observed that adding an extra target item (e.g., "other answer") could make the questionnaire more flexible and it could capture other students' interpretations.

Students' behaviour Screen and video recordings made it possible to observe the students' actions. For instance, S8 seemed to have learnt how to read the x-axis during the first question. He had problems answering question 1, but after looking at the target items, he managed to answer it correctly. However, he tried to apply the same strategy

Question type vs. Participant	S8	S9	S11
1. Reading from horizontal axis	✓	✗	✓
2. Reading from vertical axis	✗	✗	✓
3. Performing operations	✗	✗	✗
4. Interpretation	✗	✗	✓

Table 3.4: Questionnaire results: wrong answers (✗) and correct answers (✓).

for question 2, but without success:

> ... how long did Peter take to walk 3 kilometres ... ah?... five! Oh! How long did it take... three kilom*... I don't get it, it has to be five! five, five...

In the case of S9, she seemed to have had difficulties when considering two variables. Recordings of the activity showed that in questions 1, 2 and 3, S9 moved her hands over the graph showing that she was constantly reading the y-axis.

S11 answered the interpretation question confidently, but he had problems when performing operations on the graph (question 3). There are no recorded utterances that help us to understand S11's answers, but it is possible that question 3 had been more difficult for all participants as it required them to read values over the negative slope of the plotted line.

Overall, the questionnaire answers, screen recordings and video recordings provided information about the students' performance and difficulties. We noted that the different data sources complemented each other and helped us to observe important subjects' conceptions, actions and strategies. For instance, the screen recording and video/sound recordings indicated that S8 and S9 performed the following strategy for questions 1, 2 and 3: they looked at the graph, made up an answer and looked at the target items to find their answers. However, when answering the interpretation question they read one target item, looked at the graph and read the next target item.

3.1.6 Suitability of the tasks and implications

Note that it is not our objective to use any of the data presented here to make any generalisations about students' performance, but to observe whether the activities employed are useful to answer our research questions.

Learning environment

Participants enjoyed the Racing Car activity and were successful when answering interpretation questions. That is, this activity seemed to be useful for helping students to relate graph features with the corresponding concepts of the situation involved. This is encouraging since it suggests that the Racing Car activity seems adequate for primary school children. We noted, however, that a couple of aspects of the environment could be improved:

- The task needs to show more tracks. In particular, TT3 could be part of the proper activity since it helped students to convey some relevant concepts.
- Although the presentation of the car was quite abstract, the youngest participant did not seem to be very affected by it (his utterances indicated that he saw the "rectangle" as a "car"), but the oldest one indicated a preference to have a more real environment. In order to make the task more user-friendly across our target group, this issue should be addressed and the environment should have a more "real" feeling.

An interesting question to address would be to see how abstract the environment could be. Would a less embellished environment be less distractive for students? This could be an issue, which can be investigated, in future research.

Assessing students' knowledge

Assessing students' answers proved to be challenging. A simpler card-sort task might be more useful in order to obtain accurate information of students' abilities. There were three main points that needed to be addressed:

- The items need to be improved. There were different issues that could facilitate analysis and observation of students' graphical knowledge. For example, items could be more consistent within its group (e.g., the "graph" cards could only have well-formed graphs without any extraneous symbols/icons/etc.). A larger set of cards could be presented to students to give more opportunities to express their knowledge. Items could be presented in the same sequence to all students to facilitate analysis. Allowing interposing cards could improve students' experience with the task as well.
- Changing the wording of the instructions to terms such as "graphs" or "diagrams" could help to clarify instructions and to reduce ambiguity of the students' answers.
- In order to clarify instructions, students could practise sorting the cards before the proper task too – similar to what they did during the Racing Car activity.

Students' pictorial interpretation

The graph comprehension questionnaire seemed useful to trigger the WYSIWYG p-prim. However, a minor modification to the target items may be needed to make it more flexible (e.g., add "other answer" option).

Technology

This pilot has described technology-based methodologies designed for enhancing and assessing students' knowledge. The case-study evaluations demonstrated the potential of the approach.

Methodologically, data triangulation from several sources played an important role in detecting participants' graphical conceptions, reasoning strategies and observing some students' problems as well. For instance, when comparing answers from the questionnaire with screen recordings made in the card-sort, it was found that S9 had difficulties when integrating two variables and tended to look only at the y-axis values.

Using the touch-screen was also helpful as it recorded the places where students were pointing – allowing for better interpretations of their answers. Although the touch-screen worked well for large items (as the ones presented in the card-sort task), students had more problems when dragging the car in the Racing Car activity. Touch-screen devices are becoming pervasive and their inclusion in school environments should be carefully considered. In the next section we mention some of the touch-screen technologies we considered for our activities and we suggest which one seems more reliable and suitable for our tasks.

3.2 Technology development

As was noted above, we required a more responsive interactive technology for our subsequent experiments than the one used in the pilot study. In this section, we perform a summative usability study to review and compare our experiences of three different types of interactive technologies. We determine to what degree the technologies meet our requirements and establish which system is most suitable and reliable as an interactive learning technology to be used with young children.

3.2.1 Aims and requirements

We were looking for a large interactive touch-screen surface that could allow students to drag large objects around the screen (for the card-sort task) and that could effectively respond to fast/slow hand gestures (for the Racing Car activity). In order to achieve this, we needed to establish a set of minimum and main requirements that the selected technology should meet.

Requirements

We initially considered using a SmartBoard[3], but there were many difficulties with this (we only mention a few of the issues here). First, user interaction is usually through another device such as a pen, thus students would not get the kinaesthetic feedback we wanted. Second, young participants might struggle with moving objects around the screen since they might find the SmartBoard too high or too wide to move the tasks' objects comfortably and smoothly. Third, the shadow of the user (who might stand between the projector and the projected image) may pose an obstacle in the Racing Car activity: e.g., it could make the participant lose grip of the car. Fourth, not all types of SmartBoards are responsive enough when the user has to interact with very small objects; that is, they might not be sufficiently accurate to use. Although touch-screen technology keeps improving, these types of devices tend to be expensive, moving them to a different location is difficult (they are usually fixed to a wall) and not all schools have them (thus, the places where we carry out our experiments at would be limited). Considering these issues helped us determine the requirements for our technology. In particular, we needed:

- A portable touch-screen technology,
- A technology where the interface would not get occluded by the user's shadow, and
- A screen surface where students have all the objects within their hands' reach.

The three technologies we present here (Section 3.2.2) cover these minimum requirements. However, we also needed to evaluate task success, efficiency and responsiveness of the technology (see Section 3.2.5). That is, we needed:

- A technology that allows the user to complete our set of tasks,
- A technology that does not make the tasks more difficult (i.e., can the user move small objects smoothly?), and

[3] For example, https://smarttech.com

- A technology that is fast-responsive (e.g., how long does it take to see the changes on the screen after the user's hand has moved?).

An evaluation of how those technologies meet these requirements is in Section 3.2.6, but before presenting this, we give a description of the touch-screens (Section 3.2.2), participants (Section 3.2.3), procedure (Section 3.2.4) and performance metrics (Section 3.2.5).

3.2.2 Description and development of the technologies

In this section, we describe the three touch-screen technologies we evaluated (Section 3.2.6) and point out how they meet our minimum requirements.

NEC MultiSync

As mentioned before, we used a 40" NEC Touch Screen MultiSync LCD4020 display for the pilot study (see Figure 3.9a). Although it was possible to move the touch-screen between rooms, it was quite heavy and had to be carried by two people. Participants did not have to worry about covering the screen with their shadow though. The surface was large enough to permit the inspection of all the items on screen simultaneously and students were able to move objects without having to walk around.

TouchSmart

The 23" Hewlett Packard TouchSmart computer is more portable than the NEC display because it can be carried by one person. It had a large screen as well. It can be easily positioned at a height suitable for a child to use the device (see Figure 3.9b) and the screen would not be occluded by the user's shadow.

IRFT system

The infra-red finger tracking (IRFT) system is based on that described by Lee (2008)[4]. With this technology, students wore a small, high-energy infra-red LED on the nail of their index finger on their dominant hand. The LED was held in place by a small piece of medical tape. A microswitch was positioned between thumb and middle finger such that a natural prehensile action would depress the switch (interpreted by the system as a computer mouse click). Power was supplied by a battery enclosed in a wristband (Figure

[4]http://johnnylee.net/projects/wii/

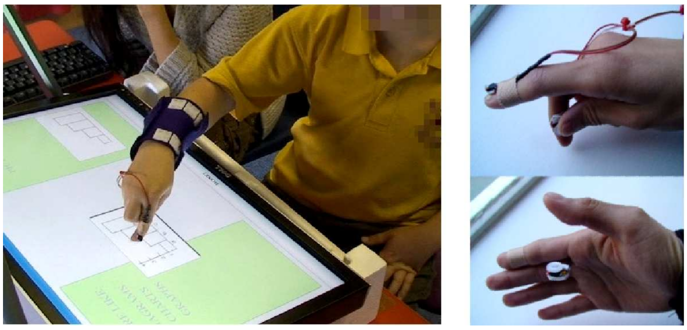


Figure 3.6: Student using the infra-red finger tracking device (left side), a close-up of the infra-red LED on a finger (top-right) and the microswitch (bottom-right).

3.6). An infra-red camera,[5] mounted about 1 metre above the students, tracked the students' finger movements as they interacted with the stimuli on a horizontally mounted LED 24" computer monitor at desk height (Figure 3.6). Lee's software[6] enabled the computer (Dell Precision T3400 Workstation) to interpret the infra-red system's input as mouse movements and clicks, via a Bluetooth wireless link.

The system was required to have the infra-red LED visible to the Wiimote remote control, otherwise the communication with the computer would break and the hand movement would not be tracked. This, however, could be avoided by positioning the Wiimote remote control at the top-front part of the participant's position. In this way, the participant would then not cover the infra-red LED from the Wiimote remote control (e.g., the reader can see the bottom part of the stand that held the Wiimote remote control in Figure 3.9c).

One of the advantages of the IRFT system was the flexibility it gave to use any surface as a touch-screen. Thus, a large 24" standard computer monitor could be used and it was even "strong" enough to withstand the pressure students put on the screen. We did not need any expensive equipment to construct the IRFT system, which cost about $50 to build. Additionally, in comparison with the technologies mentioned above, the system allowed us to position the screen either horizontally (i.e. it could lay flat on a table) or vertically (see Figure 3.9c).

Although different pieces of hardware were needed to set up the system (a PC, its monitor, a stand, and the Wiimote remote control), all parts could be easily transported from one place to another by one person.

[5]The infra-red camera of a Nintendo "Wiimote" remote control.

[6]http://johnnylee.net/projects/wii/

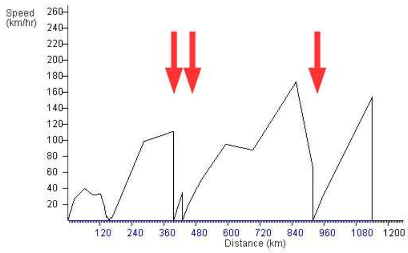


Figure 3.7: Example of how sudden stops are shown in the plotted line. The arrows point to three sudden stops.

3.2.3 Participants

We had a group of 14 students who tried the different touch-screen technologies using an improved version of our tasks (which are presented in Section 4.3). Five of those students used the TouchSmart (mean age was 10;1 years), five used the NEC display (10;8) and four used the IRFT system (8;1). In addition, if the parent and the student gave their consent, the session was videotaped and data logs were recorded.

3.2.4 Procedure

All participants followed the sequence of tasks as described in Section 4.3 and the procedure followed was as described in Section 4.4.2.

3.2.5 Performance metrics

As mentioned before (Section 3.2.1), we wanted to evaluate task success, efficiency and responsiveness of the touch-screens. We determined these aspects by:

- Observing whether students were able to finish all the set of tasks,
- Measuring any sudden stops in the graphs produced by students during the Racing Car activity, and
- Measuring the time delay (response time) between the car movement and changes in the graph.

Task success

This was recorded as the proportion of students who finished the set of tasks.

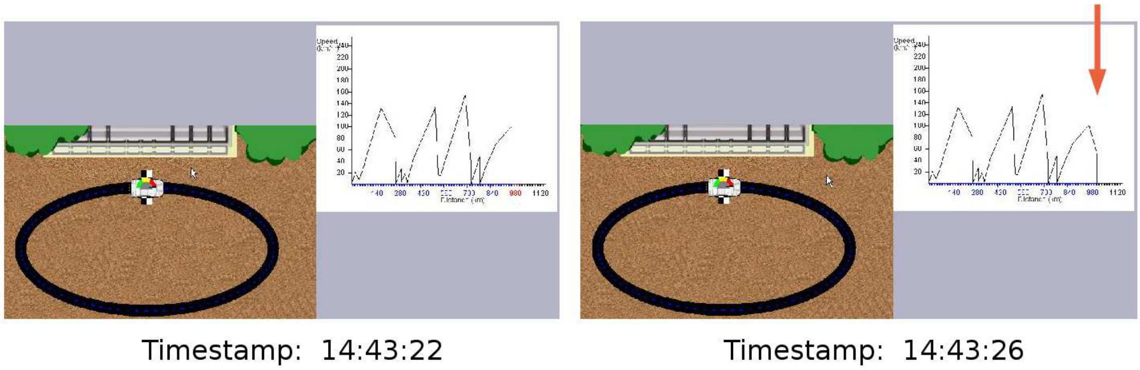


Figure 3.8: Screenshot from a race and its timestamp (minutes:seconds:frames) - there are 30 frames per second. The left image shows the timestamp of the car when it reached the end line. The right image shows the timestamp when the plotted line was fully plotted.

Efficiency

For the purpose of this study, efficiency is an indicator of how easy or difficult is for a participant to manipulate the car during the Racing Car activity. It was noted that there were several sudden stops during the race when students found it difficult to move the car. For example, the graph in Figure 3.7 shows three sudden stops, which can be observed in the plotted line as three clear vertical straight lines.

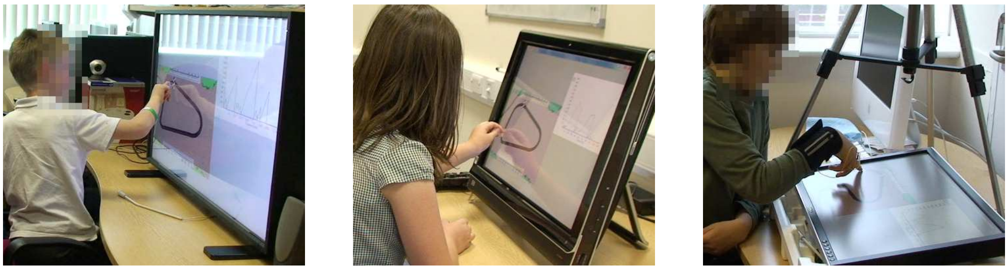
We indicated the efficiency of each technology by measuring the mean number of sudden stops encountered in the participants' graphs.

Responsiveness

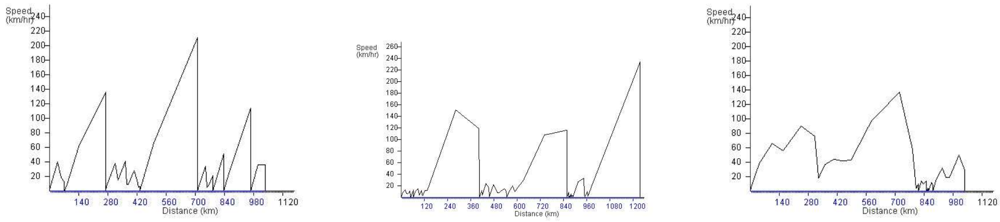
We determined how responsive the device was by looking at the difference between the time the car got to the end of the track and the time the plotted line was fully plotted. For example, the race in Figure 3.8 had a delay of 4 frames (a delay of 132 milliseconds). We used the Racing Car activity to observe the devices' responsiveness because it was clearer to point out when the car had stopped and when there was a perceived change in the display. Thus, responsiveness was measured by the mean time delay for all the participants' activity races.

3.2.6 Results: evaluation of the technologies

Our observations on how the three technologies meet our main requirements are below.



(a) NEC MultiSync. (b) TouchSmart. (c) IRFT system.



(d) Graph created using the NEC MultiSync. (e) Graph created using the TouchSmart. (f) Graph created using the IRFT system.

Figure 3.9: Video screenshots of students using the NEC MultiSync display, the TouchSmart computer and the IRFT system (top). The graphs generated by those students when racing the car on track AT4 are shown at the bottom.

NEC MultiSync

Task success All 5 students (100%) finished the set of tasks presented to them. This suggests that we can rely on this technology to administer them.

Efficiency An example of a graph generated by a student when racing the car on a track can be seen in Figure 3.9d. Students who used this display seemed to have some difficulties to control the car. This can be observed in the graph in Figure 3.10, which shows the mean number of sudden stops for each of the activity trials. From this graph, we can see that the NEC display had the highest mean of sudden stops for 4 out of the 6 activity trials. Overall, the mean number of sudden stops was 2.1 for this display.

Responsiveness Figure 3.11 shows the mean response time for each students' trial. We can see that the NEC display was the slowest to show the changes in the graph for all activity trials. The mean response time for this device was 236.3 milliseconds.

Additional observations The vertical position of the screen could make the whole set of tasks tiresome since users needed to keep their hand extended constantly (see Figure 3.9a).

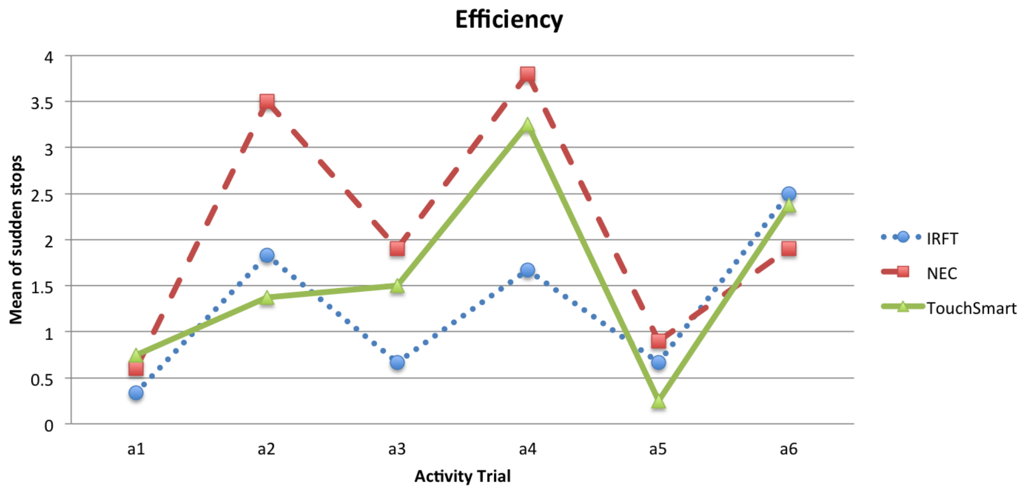


Figure 3.10: Mean number of sudden stops for each activity trial per technology.

TouchSmart

Task success We found that the TouchSmart was not sufficiently reliable: the screen "froze" or the computer crashed in two out of the overall five sessions. This meant that in those cases, (1) we had to wait for the computer to restart and setup the testing environment again while the student waited, and (2) we lost part of the data collected. Thus, only 60% of the participants were able to successfully finish all the tasks.

Efficiency It was not possible to recover data once the computer had crashed; thus the graphs that were reviewed were from 4 of the participants only. An example of a graph generated by a student when racing the car on a track is in Figure 3.9e. The graph in Figure 3.10 shows that students using the TouchSmart display might had had more difficulties to manipulate the car on a couple of the activity trials. Overall, the mean number of sudden stops was 1.6 for the TouchSmart display.

Responsiveness Interestingly, from the video recordings there were several times where we did not find a delay between the time the car stopped and the time the graph was fully plotted. This is reflected in the mean response time for each activity trial (see Figure 3.11). The overall mean response time for the TouchSmart was 85.8 milliseconds.

IRFT system

Task success All 4 participants (100%) finished the set of tasks. Thus, this interface seemed reliable to administrate all the activities.

Efficiency We did not get permission to video/screen record the interactions from one of the participants, thus the data observed was from 3 participants only. An example of a graph generated by a student when racing the car on a track is in Figure 3.9f. From the graph in Figure 3.10 we can see that students seemed to have more problems to race the car during the activity trial a6, which is considered to have a high degree of difficulty (see Section 4.3.2). However, for the rest of the activity trials, participants had a relatively low mean of sudden stops. Overall, the mean number of sudden stops was 1.3 for the IRFT system.

Responsiveness From Figure 3.11 we can see that the IRFT system has a better response time than the NEC display. Its time delay seemed very close to the one for the TouchSmart, too. The mean response time for the IRFT system was 101.2 milliseconds.

Additional observations Children (and even adults) who tried the device seemed comfortable wearing the IRFT system. One of its disadvantages, however, was that students noted it was possible to lift the finger off the screen to move the items. Nonetheless, when doing that, manipulation of the items became difficult since the movements picked up by the system were no longer in line with the previously calibrated signal. Usually, after a quick reminder to keep the finger on the screen, the participant maintained the finger on the screen from then on. A second issue was regarding the use of frames. Lee's software caused some problems when showing the cards during the card-sort task (i.e. the contents of the frame were not completely shown), but this rarely happened and it was quickly resolved by first minimising and then maximising the application, which took about 2–3 seconds to perform. Therefore, it was not much of an obstacle to doing the experiment.

3.2.7 Overview of the technologies and conclusion

An overview of the most relevant characteristics of the interactive technologies mentioned above is in Table 3.5. Overall, we were looking for touch-screens that could be used in school and laboratory settings and that were reliable, efficient and responsive, but also large and portable. Although all of the interactive technologies had disadvantages, only one seemed suitable to use for our interactive learning environment: the IRFT system. We noted that the NEC display was neither responsive nor efficient enough, thus students had problems controlling the car during the Racing Car activity. The SmartTouch showed good responsiveness and it seemed as efficient as the IRFT system; however, the risk of losing data overweighed the benefits that this technology provided (e.g., acceptable

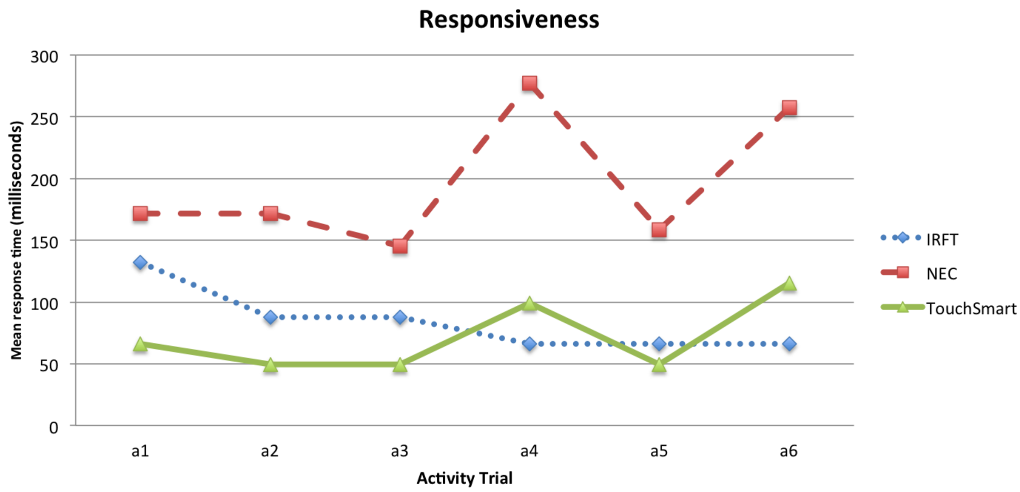


Figure 3.11: Mean response time (milliseconds) for each activity trial, per technology.

response time, portability, etc.).

	NEC	TouchSmart	IRFT system
Reliable	✓	✗	✓
Efficiency score	<u>2.1</u>	1.6	**1.3**
Responsiveness score	<u>236.3</u>	**85.8**	101.2

Table 3.5: Overview of the interactive technologies that we reviewed. The best score is in bold and the worst score is underlined.

We found that, from the 3 technologies reviewed, the IRFT system was the most adequate technology to use for our tasks. In addition, our observations could inform the selection of a touch-screen for other learning environments that have similar requirements to our system. It is important not to assume that any interactive learning technology will be suitable for learning, that is, the task has to be matched with the adequate technology. For example, at the time of this study, there were no iPads or similar devices available. Nonetheless, such devices might not have been used since the working area is small (e.g, card items would have been hard to recognise on screen).

3.3 Conclusions

We aimed to find out if the tasks we designed were suitable for primary school students. The tasks were evaluated on the basis of the feedback obtained from three students and observations from different data sources such as video and sound recordings.

Overall, we found that the tasks seemed appropriate to be used with young students, but there are some parts of the design that have to be improved; particularly regarding the instructions of the card-sort task. We also found that the IRFT system is a useful interactive technology that could be employed with our interactive learning environment. In summary:

- The Racing Car activity seemed useful for helping students to improve their line graph interpretations. However, it is suggested that we build a more child-friendly interface that shows a more realistic setting rather than an abstract one. It is recommended that we increase the number of tracks in order to allow more opportunities to explore the environment too.

- The card-sort task's instructions could be improved by clearly specifying its categories (e.g., "graphs") and/or instructions could be demonstrated too. In addition, a large set of cards could be presented to students to increase the opportunity to express their knowledge.

- The graph comprehension questionnaire might need at least an additional target item (e.g., "other answer") in order to consider other possible students' conceptions.

- The IRFT device seemed efficient, had an acceptable response time and it was reliable. It could potentially be used for other interactive learning environment systems as well.

These issues were addressed and considered for the first experiment, which is presented in the next chapter.

Chapter 4

Experiment 1

The aim of this chapter is to present an experiment that could help us to answer our two research questions: (1) which form of interactivity is more effective for improving students' interpretations of line graphs? (2) How can we assess students' interpretations of line graphs?

Improved versions of the activities presented in the pilot study were administered to 37 primary school students between the third- to sixth-year at school. Participants were assigned to one of two interventions that utilised a modified version of Janvier's [38] Racing Car activity: students either raced a car along a track (active mode) or, in a yoked condition, watched somebody else's experiments (passive mode). In addition, an enhanced version of the card-sort task was used to assess students' graphical knowledge.

The results present two main findings. First, there was some indication that the active mode could be more useful in helping students to improve their knowledge of line graphs than the passive mode. Second, we found evidence of varieties (or degrees) of pictorial interpretations (e.g., a "mixed" conception is described as a simultaneous pictorial and correct interpretation of the graph). This last finding not only has implications about how students should be assessed, but it highlights that the term "graph-as-picture misconception" is an imprecise term for describing a wide range of students' pictorial conceptions as well.

4.1 Research questions

The next two sections give a brief summary of each of our research questions and of how we are tackling them. Similar to the previous chapter, students interacted with the environment through a touch-screen in order to give an embodied experience to students.

Note that we targeted students from the 3rd to 6th school year (7-8 to 10-11 years old) since we have seen that students as young as 7-8 years old were able to understand line graphs (see Section 2.3.1 and [1], [13], [54], [19], [33]).

4.1.1 Which form of interactivity is more effective to improve students' line graph interpretations?

In the previous chapter, a Racing Car activity was implemented. The Racing Car activity was based on (1) the original paper-and-pencil Racing Car activity reported by Janvier [39], and (2) other research on interactive environments designed to support students' understanding of abstractions (e.g., [60], [70]). As was mentioned in the previous chapter, we needed to improve the learning environment presented in Section 3.1.2; this was done mainly by (1) showing a more realistic setting, and (2) adding more activity trials (see Section 3.1.6 and Section 3.3).

In addition, two modalities of the Racing Car activity were presented. In the active mode, students were allowed to "drive" a car along several racing tracks. In the passive mode, students watched a car moving along several racing tracks. In both modalities, a speed/distance graph was presented alongside, and as the car was racing or was being raced along the track, the graph showed the car's speed data concurrently. Each race was replayed so that students had the opportunity to observe the changes in the graph as well. In order to be able to compare similar experiences between the passive and the active modes, it was decided to have a yoked design for this task; that is, the activities performed by participant X in the active mode were watched by participant Y in the passive mode. As we discussed in Section 2.5.3, some studies that support the use of active modes over passive modes might not be comparing both modalities fairly (e.g., [3]); but when exposing participants (in both conditions) to similar equivalent visual information, the passive modality could be as effective (or more) as the active modality (e.g., [41]). Thus, in order to give similar experiences to both groups, we employed a yoked design.

Students were asked to interpret other line graphs after the Racing Car activity, too, for the purpose of finding out more about their conceptions and experiences.

4.1.2 How can we assess students' knowledge of line graphs?

We aimed to improve the method for assessing students' interpretations of line graphs using an enhanced version of the card-sort task presented in the previous chapter (Section 3.1.2). The card-sort task was mainly improved by (1) adding more items and (2) reducing the number of cards presented on the screen per trial (see Section 3.1.6 and Section 3.3). Note that the task also presented items showing line graphs and other representational forms, but here we discuss results regarding the line graph items only.

Accuracy scores and verbal utterances were obtained for assessment. The items presented to students showed diagrams, pictures, and pictograms (graphs conflated with pictures). More details of this task can be found in Section 4.3.1.

4.2 Hypothesis

In this experiment, we hypothesised that:

> The active mode will help 7-11 year-old students to interpret line graphs more correctly than the passive mode.

Students in the passive modality might spend more cognitive resources trying to make sense of what is happening on the graph and track. However, in the active modality, students have the advantage of testing their own hypotheses and of being able to observe their experiments during the replay. Additionally, the kinaesthetic feedback from the friction of the students' fingertips against the touch-screen and the students' gestures might enhance the students' associations of the changes in speed of the racing car and changes in the graph. As Beichner claimed *"feedback appeals to the visual and kinesthetic senses. A simple visual juxtaposition of event images and graphs is not as good as seeing (and "feeling") the actual event while graphs are being made"* (p. 812) – see Section 3.1.2, Section 2.5.2 and Section 2.5.3.

4.3 Experimental design

A diagram of the experiment design is shown in Figure 4.1. The Discrimination Task (DT or card-sort task) functions as an assessment method for the intervention. The intervention is composed of the Racing Car (RC) activity and the Select Track Questionnaire (STQ); both activities were obtained from Janvier's set of tasks, which aimed to observe students' interpretations of line graphs.

Observe that a Graph Comprehension Questionnaire (GCQ) is administered prior to the intervention to help us to screen the students who trigger the WYSIWYG p-prim more readily than the others in this task (see Section 2.4.2 and Section 2.6.2). That is, the GCQ was used to create two groups: students whose main interpretation involves pictorial answers and students whose main interpretation do not involve pictorial answers. The GCQ was administered after the pre-Discrimination Task (pre-DT); if we had given the GCQ first, we could have affected students' performance on the pre-DT activity since the students would have read and interpreted a line graph in advance of the assessment. Note that the students did not provide utterances in the GCQ and the time they spent in the GCQ was much less than the time they spent on the intervention.

Figure 4.1: Overview of the experimental design.

More details of each task are provided in the next sections (screenshots of the tasks are shown in Figure 4.2).

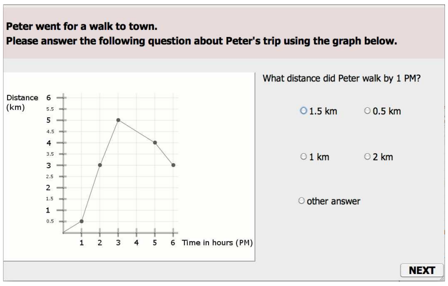
4.3.1 Observing students' interpretations: Discrimination Task

In order to know more about students' graphical knowledge, the card-sort used in Section 3.1.2 was improved (as discussed in Section 3.1.6 and Section 3.3) and a diagrammatic/picture classification task was created. In this task, we examined how students sorted pairs of cards instead of observing how they ranked three different items (which was difficult to analyse in the pilot study). The pairs showed graph/graph, graph/picture and picture/picture items. In addition, items showed a continuum of features ranging from "pure" diagrams to "pure" pictures. We expected to obtain an insight into students' knowledge from their classifications and their utterances in this task.

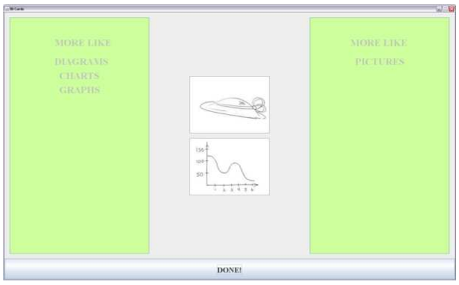
We provided two categories for the students to sort: (1) pictures, and (2) diagrams, graphs or charts. Figure 4.2b shows a screenshot of the task and Appendix B (Table B.2) shows the cards used and how they were expected to be classified.

The assessment of the Discrimination Task was done by looking at the students' scores: the students' classifications were compared with the arrangement of the cards in Appendix B, and for every item that was classified as in Table B.2, the student obtained a point. Thus, scoring was objective because it would be the same if different scorers did it.

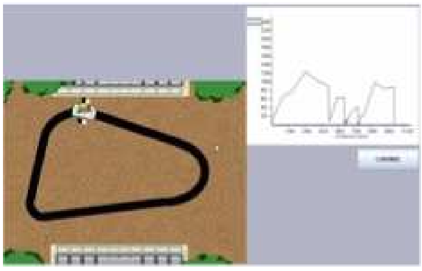
Additionally, the pre-Discrimination Task (pre-DT) and post-Discrimination Task (post-DT) were parallel forms. Although the items presented in the pre-DT and post-DT were different, care was taken to equate diagrams and pictures in terms of visual complexity and visuospatial features. That is, for each item in the pre-DT we tried to create an item



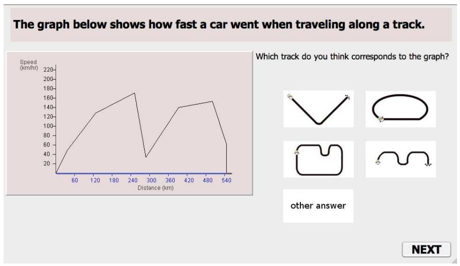
(a) Graph Comprehension Questionnaire.



(b) Discrimination Task.

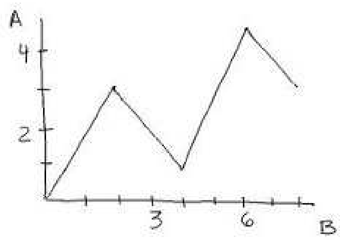


(c) Racing Car activity.

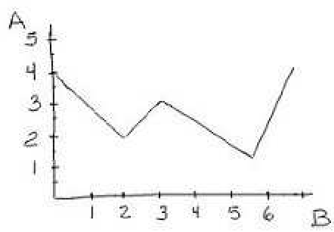


(d) Select Track Questionnaire.

Figure 4.2: Screenshots of the tasks used in experiment 1.



(a) Item in pre-Discrimination Task.



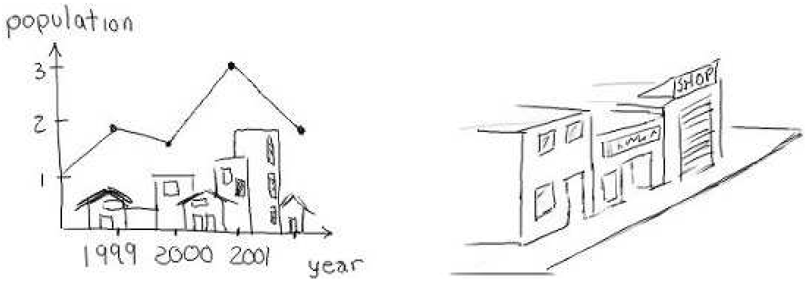
(b) Item in post-Discrimination Task.

Figure 4.3: Two parallel items in the pre-Discrimination Task and post-Discrimination Task.

in the post-DT with similar characteristics to the item in the pre-DT. For example, if an item in the pre-DT (see Figure 4.3a) showed a line graph without bends (sharp changes of trend) then its corresponding item in the post-test would show sharp bends too (see Figure 4.3b). We tried to keep similar features within pairs too. For example, if a graph item showed pictures of buildings then its picture item showed pictures of buildings (see Figure 4.4).

We tried to vary examples in order to avoid having the pre-DT and post-DT exactly the same; in other words, we wanted to reduce the possibility of correct answers due to practice. There were several iterations to improve the card set (cards were simplified, made clearer, etc.) until it was considered to be similar enough.

As we mentioned before (see Section 3.1.6 and Section 3.3), we found difficulties as-



(a) Graph. (b) Picture.

Figure 4.4: One pair of the pre-Discrimination Task.

sessing how students organised 3 items; hence, we reduced the number of cards presented to students. We kept a couple of items on the screen to observe if students made any type of comparisons between them and, if they did, observe which characteristics of the items might have helped them to make their decisions. For example, utterances like these could give some indication of the students' understanding of the graph: *"this one has numbers then it goes in the diagrams"* or *"this has pictures and goes in pictures, but this one does not have pictures so it goes in diagrams"* or *"this shows how much you pay for something,"* etc.

Material

A total of 24 line-graph items were used in the pre-DT and another 24 line-graph items were used in the post-DT[1]. Each item consisted of hand-drawn monochrome JPEG images of diagrams (50%) and pictures (50%). The diagrams were based on published primary school resources (e.g., UK National Curriculum material, teaching resource books, etc.).

Items were presented in 12 unique pairs. There were six forms of the pairs: (1) two pictures, (2) two "pure" diagrams, (3) a diagram with pictures (pictogram) and a picture, (4) a "pure" diagram and a picture, (5) a diagram and a picture that preserves the structure of the diagram, and (6) a diagram and a picture created by a small modification of the diagram. Examples and descriptions of these categories are shown in Table 4.1. Note that the pairs were designed to represent a continuum: from "pure" diagrams to "pure" pictures.

Pairs were shown in the same order to all participants in the pre-DT. However, in order to reduce the possibility of students getting correct answers due to practise, the equivalent items in the post-DT were shuffled and presented in that (new) order to all participants.

[1]There were 144 items of tables, bar charts, line graphs, pie charts, hierarchies/network diagrams, and set diagrams, but we consider the line-graph items in this thesis only.

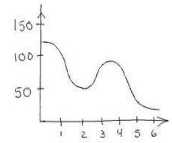
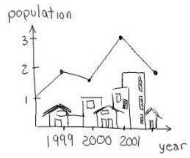
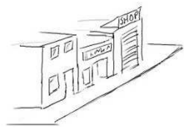
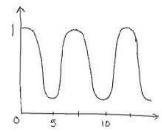
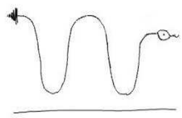
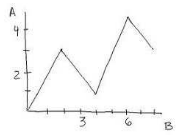
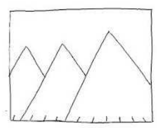
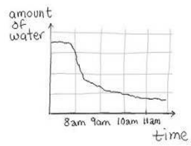
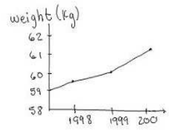
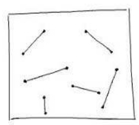
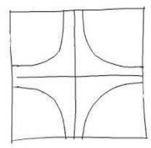
Pair-type description	Item 1	Item 2
A diagram and a picture.		
A diagram with pictures (pictogram) and a picture.		
A diagram and a picture that preserves the structure of the diagram.		
A diagram and a picture formed from a small modification of the diagram.		
Two diagrams.		
Two pictures.		

Table 4.1: Examples of stimulus pairs presented in the Discrimination Task.

4.3.2 Intervention: Racing Car and Select Track Questionnaire

The intervention was formed by the two tasks Janvier originally presented to their students: a Racing Car activity and a Select Track Questionnaire. The aim of these tasks was to *"analyse the disrupting effect of co-ordinating two modes of representation (graph and picture) in the handling of many global features of the graphs"* ([38], p. A3.4). Our own version of these activities is presented here.

Racing Car activity

This activity was an enhanced version from the one presented in the previous chapter (Section 3.1.2). The interface was improved and more tracks were added to the activity.

Also, the instructions changed with the purpose of showing the task “by example”.

There were two intervention modes in this study. In one mode, the student “drove” the racing car along a track and a speed/distance graph was plotted concurrently alongside (active mode). In the other mode (passive mode), the participants watched one of their peers’ experiments (see Section 4.1.1 and Section 2.5.3). In the passive mode, students could devote their attentional resources to observing the races and the corresponding changes in the graph; whereas in the active mode, students could test their own hypotheses during the task. Hence, having the two conditions allowed a comparison of learning outcomes from “observing” versus “doing” (or “passive” versus “active”, see Section 2.5.2).

Students were allowed to race (active mode) or to watch (passive mode) a car along six different tracks. Tracks varied in difficulty and form; some of them were borrowed from examples of Janvier’s task ([38], [39]), see Table 4.2. The Racing Car activity was preceded by a trial session composed of six other tracks. Three of the trial tracks were used by the experimenter to demonstrate the activity and the three other identical tracks were used by pupils to familiarise themselves with the task (see Figure 4.7). The experimenter and the pupil alternated trials too. Instructions given to participants are in Appendix B (Figure B.1), and a screenshot of the activity is in Figure 4.2c.

Racing Car tracks The tracks for the activity are presented in Table 4.2. Note that the track number indicates the order in which the tracks were presented to students. Open tracks (numbers 1, 3 and 5) have different “start” and “end” points, whereas closed tracks (numbers 2, 4 and 6) start and end at the same place. It can be considered that the order of the tracks in Table 4.2 (from left to right) reflects their degree of difficulty. This can be defined by the number of bends in the track and by the difficulty of racing the car on tracks that have sharper bends.

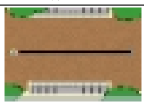
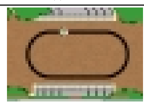
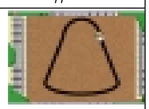
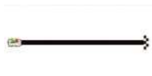
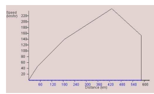
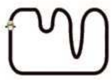
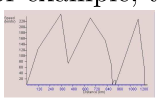
Open tracks			Closed tracks		
#1	#3	#5	#2	#4	#6
					

Table 4.2: Tracks used in the Racing Car activity by degree of difficulty (difficulty increases from left to right). The number on top of the track indicates the order in which the track was presented to students.

Select Track Questionnaire

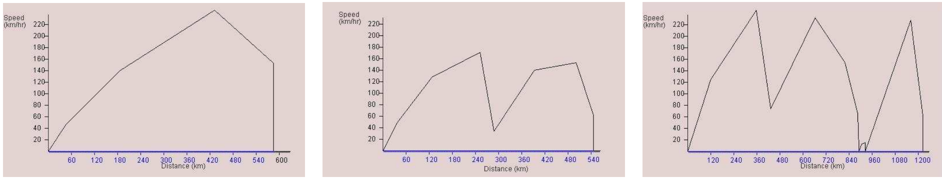
In order to observe the students' coordination of the graph and track, they were asked to actively provide information about their interpretations of different line graphs. This task was particularly important in developing a task analysis, which helped us to model how students could have solved the task (i.e. how the graph could have been interpreted) and how pictorial interpretations could have occurred (see Section 4.5.4). The Select Track Questionnaire presented four different questions (see Appendix B, Table B.3). A graph that showed the journey of a car was presented on the left hand side of the screen, while on the right hand side of the screen, the question and its target items were shown (see Figure 4.2d). In this task, the student had to select the track that might have created the graph presented. For each question, there were five target items (including one that read "other answer"). Target items were selected while considering the following criteria:

- One item (the correct answer) was the track used to generate the graph. For example, a smooth race on a track like this  generated the graph .
- One item (the pictorial answer) resembled the plotted line and the horizontal axis of the graph. For example, the track  resembles the plotted line and x-axis of this graph 

In addition, the questions' target items were constructed considering the next description:

- Question 1 showed tracks that were presented to students during the Racing Car activity.
- Question 2 showed two tracks that were presented to students during the Racing Car activity and two tracks that were not presented to them before.
- Question 4 showed tracks that were not seen by students during the Racing Car activity.

Question 3 differed from the rest; instead of giving track-options as target items it presented text-options as target items. It also asked students to interpret a graph created by a race on a track they have not seen; that is, it asked students to infer how the track could have been (in particular, it asked how many bends the track could have had).



(a) Question 1 (b) Question 3 and 4 (c) Question 2

Figure 4.5: Graphs used in the Select Track Questionnaire.

The graphs used in the Select Track Questionnaire are in Figure 4.5. It can be observed that they differ in complexity, too. Examples of the tracks presented for Question 2 are in Figure 4.6 (on the left two “known” tracks, and on the right two “unknown” tracks).

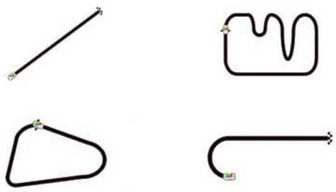


Figure 4.6: Example tracks that correspond to question #2 of the Select Track Questionnaire. See Figure 4.5c.

4.3.3 Graph Comprehension Questionnaire

As we discussed in the previous chapter (Section 3.1.2), one of the common means to observe students’ pictorial interpretations is through a multiple-choice test (e.g., [44], [70]). Thus, we used this method to help us screen out students who were giving pictorial interpretations. The questionnaires presented by Curcio and Smith-Burke [26] and Janvier [39] were used as a reference for the one presented here – the former research provided examples of “situations” that could be used as a background story and the latter provided the general structure of the questionnaire. As before, our Graph Comprehension Questionnaire (GCQ) covered three typical aspects of students’ evaluations on line graphs: reading axis (“y” and “x”), perform operations, and interpreting the graph (see Section 3.1.2). A gender-neutral cover story topic for the scenario was chosen (a child going for a walk) in order to avoid any gender bias answers as in Janvier [39].

In response to the participants’ comments during the pilot study (Section 3.1.2), few changes were done to the GCQ. This means that the target items were improved and they included “other option” as a target item. In addition, participants were instructed to reflect on the line graph just before answering the final question:

> Think about this graph. Take your time. When you feel ready to answer the last question press this button [NEXT].

This request was presented on an additional screen, which was presented before the last question to prepare students for the interpretation question. Details of the GCQ are in Appendix B, Table B.1 and a screenshot of this task is in Figure 4.2a.

4.3.4 The pictorial group

Question number four of the GCQ was of great interest for us as it presented different interpretation options. There were two target items that were related to pictorial-like responses. One target item (*"He walks up a hill and then goes down the other side"*) presented one of the prototypical forms of a graph-as-picture interpretation. This option suggested that none of the parts of the representation (the graph) were interpreted abstractly and part of its features (the plotted line) were being mapped directly to an object related to the situation (a hill). A second target item (*"He walked fast and then he went down a bridge"*) had two parts: the statement started with a partial and correct interpretation of the graph (*"He walked fast"*) and it was followed by an non-abstract interpretation of the graph (*"he went down a bridge"*), which can be considered as a pictorial interpretation. In both cases, (part of) the graph is interpreted as a picture; thus, we could say that students who selected any of those options had cued the WYSIWYG p-prim more strongly than the other students (see Section 2.4.2).

Note that these two distracter options were used to create two groups: (1) students who chose any of the pictorial options (the pictorial group) and (2) students who chose a non-pictorial option.

The objective of this was to observe how students who initially answered with a pictorial answer perform after the intervention. Note that this was not another independent variable of the experiment, but it was a way to observe the effect of our two modalities for a sub-group of the participants (the pictorial group).

4.4 Method

4.4.1 Participants

A total of 37 students participated in this experiment, but data from 6 students had to be excluded since they did not perform or finish some of the tasks. Thus, data from 18 female and 13 male students was analysed. There were 8 students in the third school year ($M =$ 7;1 years), 8 students in the fourth school year (8;10), 10 students in the fifth school year (9;7) and 5 students in the sixth school year (10;8). All participants attended a public (state) elementary school.

A schoolteacher invited students from different grades to participate in the study. The students who had parental consent to participate were randomly assigned to one of two groups: the active or passive modes of the learning environment. Thus, it could be considered that both groups had a representative sample of students with different skills. In addition, an attempt was made to assign equal numbers of participants of each school year to each mode.

4.4.2 Procedure

Participants were tested individually (and as they were available) in a room next to two classrooms. Where consent was given, sessions were videotaped. The computer screen activity was dynamically recorded[2] and interactions with the software were logged. All participants performed the task using an infra-red finger tracking (IRFT) system as well (see Section 3.2.2).

Initially, participants were introduced to the touch-screen device and shown how to use it. Students then spent a few minutes drawing with a painting application in order to familiarise themselves with the interactive system.

The sequence of tasks is described in Figure 4.1. The instructions for the activities were as explained below.

Discrimination Task (pre-test)

The instructions were a short introduction to the activity since the experimenter demonstrated this task. The students were asked if they preferred to read the instructions by themselves or to have them read by the experimenter. The instructions were:

> In each of the next screens you will see some cards that need to be sorted in 2 groups: (1) more like diagrams, charts, graphs; (2) more like pictures. You have 5 minutes to drag as many cards as you can into the areas: "more like diagrams, charts, graphs" or "more like pictures" and try to speak aloud whatever that comes into your mind. Can you sort all the cards before running out of time?

This was followed by a 40-second demonstration of the activity where the experimenter showed the student two screen areas for sorted stimuli. The experimenter pointed to the

[2]Using Camtasia Studio 6.0.1., Techsmith Ltd.

left side and read “diagrams, charts, graphs” and then pointed to the right side and read “pictures”. Afterwards, the two stimuli in the middle of the screen were pointed out and it was explained that those items were the cards. In the first demonstration trial the experimenter (correctly) sorted the two images: one onto the diagrams’ side and the other onto the pictures’ side. In the second demonstration trial the experimenter (correctly) sorted the two stimuli (this time both were pictures) by putting both of them onto the “pictures” screen area. After the experimenter finished the demonstration, the student was asked to sort the same sequence of cards. The student then performed the proper task and continued until 5 minutes elapsed or he or she had finished all the trials.

Graph Comprehension Questionnaire

During the task, the students read the instructions by themselves or the experimenter read them aloud:

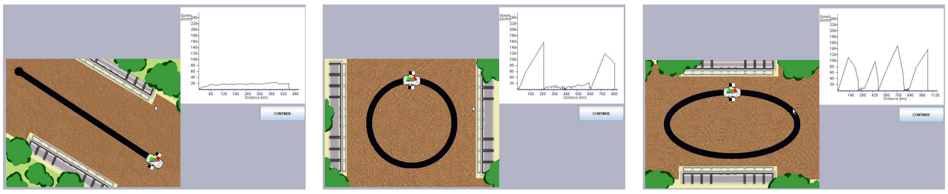
> Peter is a kid that lives in a very, very flat town. He likes to walk and he decided to go for a walk to town. The next screens show a graph about Peter’s trip. Read the questions and answer them using the graph. You can take as much time as you need.

Afterwards, the participants were told that they could begin the activity. They pressed the button labelled “next” and the first question was presented on a new screen. The experimenter read the questions if asked, and no feedback was given to the student. The task finished when the four questions were answered.

Racing Car activity

Instructions were presented in three different screens (see Appendix B, Figure B.1) and were read by the student or by the experimenter. After reading the instructions, participants were given three trial tracks to play with for the active mode. For each of the trial tracks the experimenter first gave an example of what the student was required to do and this was followed by a trial where the student was allowed to race the car at his/her will. The experimenter showed the same sequence of trial experiments for all the students; trials 1, 2 and 3 are presented in Figure 4.7a, Figure 4.7b and Figure 4.7c, respectively. Examples of the utterances and graphs produced by the experimenter are in Figure 4.7.

Instructions for the passive mode were similar to the instructions given in the active mode, but they were slightly modified to reflect the task-interaction mode. For instance,



(a) *I will go at an even pace all the time.*

(b) *Sometimes I will go fast, then slow, then fast, and slow again.*

(c) *I will go fast in the easiest parts and slow in the more difficult ones.*

Figure 4.7: Example of the trials and utterances produced by the experimenter.

during trial 1, the experimenter could have uttered *"I think it will go at an even pace all the time"* for the passive mode, but uttered *"I will go at an even pace all the time"* for the active mode.

During the session, pupils were questioned about their plans and observations and they were encouraged to "think aloud" during the task.

Select Track Questionnaire

After the Racing Car activity students were questioned on their interpretations of specific graphs. Again, they were first asked if they wanted to read the instructions or if the experimenter should read them for them. For each question, after the students had selected their answer, the experimenter asked them *"can you tell me why you selected that?"*

Discrimination Task (post-test)

The post-Discrimination Task was administered after the Select Track Questionnaire. The procedure mentioned above for the discrimination task was repeated.

Additional notes

The testing room had two doors, and school regulations required to have at least one door open all the time during the sessions. Thus, the task was interrupted on a few occasions and noise from the surrounding classrooms was heard most of the time. Although some students were able to work with this level of noise, others seemed to be slightly distracted by it.

The IRFT device did not work well on a few occasions. The Wiimote Whiteboard software had some problems when it was used with software that used multiple frames. Thus, some trials of the discrimination task did not show the items. However, this was quickly solved by a "minimising and maximising" screen action.

At the end of the session each child received a small educational gift to thank them for their participation.

4.4.3 Independent and dependent variables

Independent variables

There was one between-subjects factor: *task-interaction mode* (or *mode*).

- The *task-interaction mode* had two levels: active and passive.

There was only one within-subject factor: *time*.

- For *time*, we had two levels: pre-test and post-test.

Dependent variables

The dependent variable is the students' Discrimination Task score; see Section 4.5.2.

4.5 Results

The results are presented in two main parts. The first part is composed of Section 4.5.2 and Section 4.5.3. In Section 4.5.2, we attempted to answer our first research question by comparing the results of the Discrimination Task between the active and the passive task-interaction modes. While in Section 4.5.3, we made an effort to explain the results from Section 4.5.2 by analysing students' strategies during the Racing Car activity.

The second part has two sections, too (Section 4.5.4 and Section 4.5.5). First, we review the Discrimination Task as an assessment method of the students' interpretations (Section 4.5.4). Afterwards, in order to find out more about how students were interpreting line graphs and how their pictorial interpretations might have occurred, we conducted a task analysis of students' interpretations, too (Section 4.5.5). From this analysis, we were able to highlight some interesting instances of students' graphical knowledge.

Before presenting our results, data regarding our groups is shown first.

4.5.1 Groups of participants

As mentioned before, the Graph Comprehension Questionnaire was used as a tool to determine our groups. Two different groups were considered: children who answered the interpretation question using any of the pictorial options (pictorial group) and children who did not choose any pictorial answers (non-pictorial group); see Section 4.3.4.

	Pictorial group	Non-pictorial group	Total
Active	3	9	12
Passive	4	15	19
Total	7	24	31

Table 4.3: Frequency of students for the active and passive modalities.

The frequency of students by groups and task-interaction mode are shown in Table 4.3. In order to find out if students selected a pictorial option by chance, a goodness-of-fit χ^2 test was conducted. Results show that the number of students selecting the pictorial options differed significantly from that expected [$\chi^2 = 3.91$, $df = 1$, $p = 0.05$]. Therefore, the pictorial group favourably represented students giving a pictorial interpretation predominantly.

4.5.2 Active mode vs. passive mode: Discrimination Task

In this section we analyse the results of the Discrimination Task (DT) in order to answer our first research question: which mode would be more effective for improving students' interpretations?

Scoring

Each trial in the DT consists of two cards. A point was allocated for each card classified correctly. A maximum of two points and a minimum of zero points could be obtained per trial. Each participant obtained a mean score that was calculated by dividing the total number of points between the number of total cards seen. For example, if a student obtained 18 points and completed 10 trials (i.e., 20 pairs) then the mean score is of $18/20 = 0.9$, which is interpreted as 90% of the cards were correctly classified. Therefore, students' scores could range from 0 to 1.

Students who completed more trials did not tend to score higher than students who completed fewer trials (the correlation between the number of trials completed and the mean scores was low and was not statistically significant, $r = 0.06$).

General performance

The mean number of trials performed on the pre-Discrimination Task was $M = 13.6$, $SD = 4.7$; and on the post-Discrimination Task was $M = 15.6$, $SD = 5.1$. On average, students improved their scores significantly from pre-test ($M = 0.90$, $SD = 0.09$) to post-test ($M = 0.95$, SD $= 0.06$), $t(30) = -2.97$, $p < 0.01$, $\eta^2 = 0.23$.

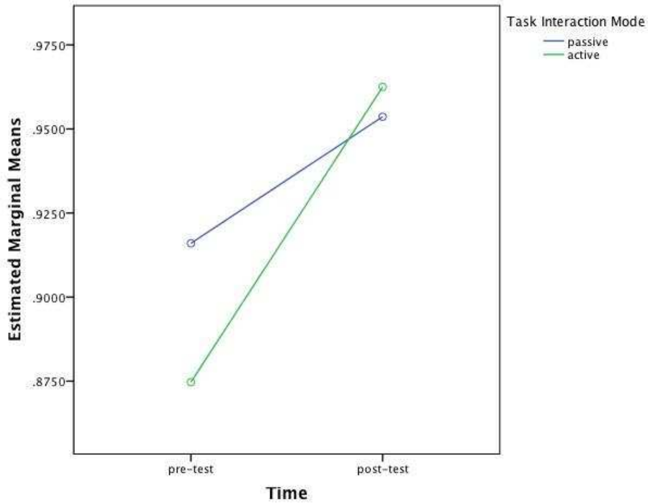


Figure 4.8: Mean scores for the Discrimination Task across the two task-interaction modes for the pre-test and post-test: all students.

Active vs. passive modes: all students

A mixed between-within 2×2 (task-interaction mode [active, passive] × time [pre-test, post-test]) subjects analysis of variance was performed on students' Discrimination Task scores to evaluate if the intervention had an effect on their performance. Figure 4.8 presents mean scores for the active ($N = 12$) and passive ($N = 19$) task-interaction modes for the pre-test and post-test.

There was a significant main effect on time, Wilks' Lambda = 0.74, $F(1{,}29) = 10.34$, $p < 0.01$, multivariate partial $\eta^2 = 0.26$; but there was not a main between-groups effect nor an interaction effect. However, taking a closer look at the results, t-tests conducted for the active mode and the passive mode showed a significant difference in the performance for the active mode [$t(11) = -2.36$, $p = 0.038$, $\eta^2 = 0.34$], but not for the passive mode. Note that an independent-samples t-test on the pre-test scores for the active and passive groups did not show a significant difference between those groups.

Active vs. passive modes: pictorial group only

A mixed between-within 2×2 (task-interaction mode [active, passive] × time [pre-test, post-test]) subjects analysis of variance was performed on the pictorial group's scores (see Figure 4.9). There was not a significant within-groups effect nor between-groups effect nor an interaction effect. However, the groups for the active ($N = 3$) and passive ($N = 4$) modes were small.

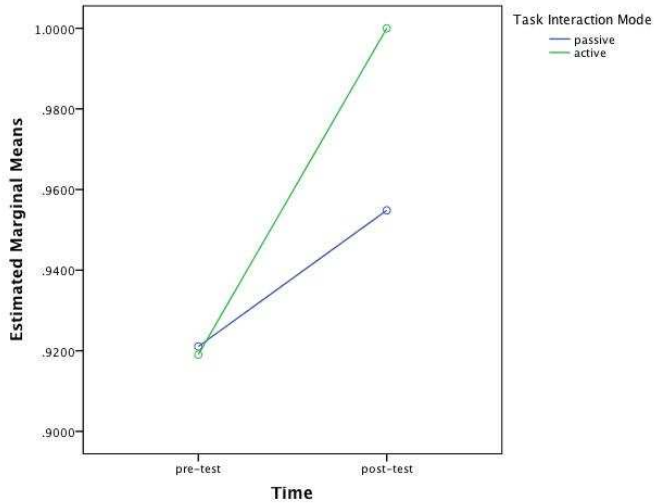


Figure 4.9: Mean scores for the Discrimination Task across the two task-interaction modes for the pre-test and post-test: students in the pictorial group only.

Results summary

We used the Discrimination Task to observe which task-interaction mode could be better to improve students' graphical interpretations. Analyses showed that the active mode might have facilitated students' interpretations, but there was not enough evidence that could show whether students in the pictorial group benefitted from the active modality as well (although the number of participants in the pictorial group was small). Note that these results need to be treated with caution since students' scores were high; thus it is possible that a ceiling effect might have happened (Figure 4.8 and in Figure 4.9 show that the mean scores for the post-test are between 0.94 and 1).

However, as a minimum, it is worthwhile seeing what students were doing in the active modality. That is, what kind of strategies students in the active modality used that could have helped them to perform better than the students in the passive modality? The next section investigates this issue further.

4.5.3 Students' strategies during the Racing Car activity

We observed that the active modality seemed to have improved students' scores. Although post-test scores were close to the ceiling, it is worthwhile reviewing what kind of strategies students were using; particularly those students that went through the active mode.

A comparison of students' strategies for both modalities was planned. However, most of the students in the passive mode produced none or a few utterances. This meant that we had to confine our analysis to a review of students' strategies from the active mode only.

Thus, in this section, we closely study students' experiences to get an insight of why the active mode might have been more effective. We coded (1) students' plans, (2) students' "focus of attention" and (3) plans' "mode". Coding details are explained next.

Procedure for data collection and the coding scheme

All students' utterances for the active group were first transcribed and only verbalisations from the activity tracks were considered. In order to observe students' behaviours during the active mode of the Racing Car, we worked on a coding scheme for: students' plans (what did the student try to do? E.g., *"I will go fast"*), students' evaluations of the race (did the student achieve what he/she wanted? E.g., *"it went very slow!"*), students' interpretations (did the student interpret the graph correctly? E.g., *"up, down, slow down a bit"*), and whether students related to the graph and track (did the student engage and relate to both representations? E.g., *"there are three {student points to the track} and there are three sides there {student points to the graph}"*). However, there was not enough information from the students' utterances to answer all those questions, except for the question related to the students' plans.

Thus, we created a coding scheme to identify what the student tried to do during the race. In addition, students' plans were categorised in order to review if there were certain patterns that could give us an insight into why the active mode seemed more successful. In particular, we focused on:

- The **mode** of the plan. The idea behind this was to observe if the student seemed to be making experiments across/within the trial (was the student racing the car at different speeds during the race?).

- The **focus** of the race. The purpose of this was to observe if the student was making his/her plans according to certain elements of the interface (e.g., shape of the track, certain form of the graph, particular colour of the speedometer, etc.).

An overview of the codes is below.

Students' plans

Some of the students' plans more commonly used were coded as follows:

- `plan_goFast`: the student tries to move the car quickly. Examples of students' utterances are: *"go fast"*, *"go speedy"*, *"go red"*.

- `plan_goSlow`: the student tries to move the car slowly. Examples of students' utterances are: *"go slow", "go green but not yellow"*.

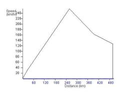
- `plan_goContinuous`: the student moves the car at the same speed. An example of students' utterances is: *"I'll stay in green"*.

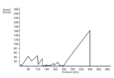
- `plan_idealCurve`: the student plans to go slowly on curves and quickly on straight segments. An example of students' utterances is: *"trying go fast in a straight and green for the corners"*.

- `plan_other`: the students' plan is not any of the four plans above. An example of students' utterances is: *"I'm making it look weird"*. Part of the student's utterance might be added to this code in order to give more information about the pupil's plan.

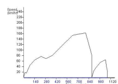
- `plan_unknown`: the students' plan is unknown due to lack of utterances and/or clear hand gestures (see below).

Note that the plan was determined either by the student's utterance or, in some cases, by the student's hand movement if no utterance was produced. In order to validate the observed hand movement, the plotted output was considered as well. For instance:

- We assigned the code `plan_goFast` for a graph that looks like  and the hand movement was "fast all the way".

- We assigned the code `plan_other` for a graph that looks like  and the hand movement looked like "fast, slow, fast".

There was an exception to this convention; if it was observed that the student got stuck *only once* (i.e. if the student lost their grip of the car involuntarily only once) and the student continued with the previous behaviour (e.g., going fast) then the plan assigned was the one he/she was previously showing (e.g., "going fast all the way"). Only one involuntary stop (that is, one "interaction error") was considered otherwise the student's plan becomes unclear. For example:

- We assigned the code `plan_goFast` if the student's hand movement was "fast", but got stuck once and the graph looked like: 

A plan coded using the hand gesture included the sub-code `hag` (hand and graph) within the code (e.g., `hag:plan_goFast`). In addition, we placed a question mark at the end of the code if there was any ambiguity when the plan was coded.

A summary of the participants' plans is in Table 4.4 and in Table 4.5. Note that the students' comments were organised by the tracks' degree of difficulty (see Section 4.3.2), but the actual order presented in the Racing Car activity is marked on the top of the tracks' picture (track number).

The student's plan (first line for each cell in Table 4.4 and Table 4.5) is either summarised by the codes described above or marked in italics when it is taken from the students' utterances. For instance, if the student stated *"I was trying to go fast"* then the student's verbalisation is summarised with the code `plan_goFast`.

Plans' characteristics: plans' mode

Each cell in Table 4.4 and Table 4.5 presents the *mode* assigned to the particular plan. There were three modes of plans: **varied**, **fixed** and **other**. A plan was marked as `fixed` when the student's aim was to drive the car at a specific speed along all the track (e.g., fast only, or slow only, or constant speed only). For example, if the student stated that the plan was to *"get this high"* or to make the *"diagram all the way up"* or to *"go fast"* then the plan was marked as `fixed`. A plan was marked as `varied` when the plan implied one or more changes of speed. For instance, if the student said go *"between fast and slow, not yellow"* or the observed (and confirmed) hand movement was "fast, slow, fast" then the plan mode was coded as `varied`.

The mode `other` was for the rest of the cases that were a mixture of different behaviours or unknown situations. For instance, some students did not have a plan in mind and moved the car without any pattern (e.g., p13 moves the car without trying anything in particular, track #6) and others expressed "feelings" about how to race the car (e.g., students could race the car using their "instinct"). That is, the `other` mode was assigned when the student's plan was not `fixed` or `varied`.

Plans' characteristics: students' focus

Students might have planned their experiences according to some features of the interface. For example, in accordance with the graph outputs (e.g., *"connect two mountains"*), or with respect to the track features (e.g., *"going fast in curves"*), or they might have expressed a pattern in terms of speed (e.g., *"fast, stop, fast, stop, four times*). In other

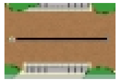
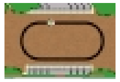
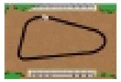
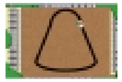
		#1 	#3 	#5 	#2 	#4 	#6 
p13	plan:	*"diagram all the way up"*	*"as fast as possible"*	**hag**: `plan_goFast`	**hag**: (fast, get stuck, fast)?	**hag**: fast, slow, get stuck, fast	not trying anything
	focus:	**drawing graph**	**speed description**	**unknown**	**unknown**	**unknown**	**unknown**
	mode:	**fixed**	**fixed**	**fixed**	**fixed**	**varied**	**other**
p14	plan:	how she felt	**hag**: `plan_goContinuous`	**hag**: `plan_goContinuous`	**hag**: `plan_goFast`	**hag**: `plan_goFast`	**hag**: `plan_goFast`
	focus:	**unknown**	**unknown**	**unknown**	**unknown**	**unknown**	**unknown**
	mode:	**other**	**fixed**	**fixed**	**fixed**	**fixed**	**fixed**
p15	plan:	`plan_goFast`	*"done a little bit and then fast"*	`plan_goFast`	`plan_goFast`	`plan_goFast`	`plan_unknown`
	focus:	**speed description**	**speed description**	**speed description**	**speed description**	**speed description**	**unknown**
	mode:	**fixed**	**varied**	**fixed**	**fixed**	**fixed**	**other**
p17	plan:	*"get this high"*	**hag**: `plan_goFast`	**hag**: fast, stop, fast-stop quickly several times	`plan_goSlow`	*"making it look weird"*	*"make mountains"*
	focus:	**drawing graph**	**unknown**	**unknown**	**speed description**	**drawing graph**	**drawing graph**
	mode:	**fixed**	**fixed**	**varied**	**fixed**	**varied**	**varied**
p20	plan:	*"really fast"*	*"slow in the middle... hit yellow and then hit red"*	`plan_goSlow`	`plan_idealCurve`	*"tried to stay in yellow to see if I could slow down... go on green... or I could go really fast"*	*"go fast... without stopping"*
	focus:	**speed description**	**speed description + speedometer**	**speed description**	**track form**	**speed description + speedometer**	**track form?**
	mode:	**fixed**	**varied**	**fixed**	**varied**	**varied**	**fixed**
p21	plan:	`plan_goFast`	**hag**: `plan_goSlow`	*"slow and then medium"*	*"slow and... medium"*	*"fast and then medium"*	*"medium"*
	focus:	**speed description**	**unknown**	**speed description**	**speed description**	**speed description**	**speed description**
	mode:	**fixed**	**fixed**	**varied**	**varied**	**varied**	**fixed**

Table 4.4: Strategies identified in the active modality. Each cell describes the student's "plan" for the trial, the student's "focus" (track, graph or speed) and the general description of the type of experience ("mode").

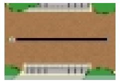
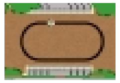
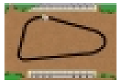
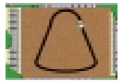
		#1 	#3 	#5 	#2 	#4 	#6 
p27	plan:	*"really fast"*	*"slow, a bit fast and really fast"*	*"slow then really fast"*	*"a bit fast and really fast"*	*"a bit fast and then I went around the bend and then a bit fast"*	plan_idealCurve
	focus:	**speed description**	**speed description**	**speed description**	**speed description**	**track form**	**track form**
	mode:	**fixed**	**varied**	**varied**	**varied**	**varied**	**varied**
p28	plan:	`plan_goFast`	`plan_goFast`	fast, under forty	`plan_idealCurve`	`plan_goFast`	`plan_idealCurve`
	focus:	**speed description**	**speed description**	**drawing graph**	**track form**	**speed description**	**track form**
	mode:	**fixed**	**fixed**	**fixed**	**varied**	**fixed**	**varied**
p29	plan:	`plan_goFast`	*"really slow and then fast"*	`plan_goFast`	*"slow then faster"*	*"medium and faster and slow and really fast"*	*"really slow and then faster... on the bend really fast"*
	focus:	**speed description**	**speed description**	**speed description**	**speed description**	**speed description**	**track form**
	mode:	**fixed**	**varied**	**fixed**	**varied**	**varied**	**varied**
p33	plan:	`plan_goFast`	hag: `plan_goFast`	hag: `plan_goFast`	*"slow then really fast then again slow"*	hag: slow, a bit fast, fast (*"yea! yellow!"*)	*"I was trying to do in there [graph]"*
	focus:	**speed description**	**unknown**	**unknown**	**speed description**	**speedometer**	**drawing graph**
	mode:	**fixed**	**fixed**	**fixed**	**varied**	**varied**	**other**
p42	plan:	hag: `plan_goFast`	hag: `plan_goFast`	hag: `plan_goFast`	`plan_goSlow`	hag: `plan_goFast`	`plan_goFast`
	focus:	**unknown**	**unknown**	**unknown**	**speed description**	**unknown**	**speed description**
	mode:	**fixed**	**fixed**	**fixed**	**fixed**	**fixed**	**fixed**
p45	plan:	*"a bit fast but... stay in the green"*	*"a bit fast but not that really fast"*	*"trying to get in the middle"*	*"trying to get a bit fast... to go a bit in the middle"*	*"a little bit fast, go in the middle, to green and then yellow"*	*"trying to keep at the same point... on the end yellow"*
	focus:	**speed description + speedometer**	**speed description**	**speedometer**	**speed description + speedometer**	**speed description + speedometer**	**speedometer**
	mode:	**fixed**	**fixed**	**fixed**	**fixed**	**varied**	**fixed**

Table 4.5: Strategies identified in the active modality. Each cell describes the student's "plan" for the trial, the student's "focus" (track, graph or speed) and the general description of the type of experience ("mode").

words, students' utterances reflected a different *focus* of attention for their plans. This is of particular interest because it could give us an idea of possible factors that could have made a difference between modalities. The codes used are listed below:

- `drawing graph`,
- `track form`,
- `speed description`,
- `speedometer` and
- `unknown`.

When the students were addressing their attention to the graph output it was said the students were **drawing a graph**; for example, a student might have said *"I want to connect two mountains"* or *"I want to go fast, under forty"*. If the student was making a plan based on features in the track (e.g., bends) then the focus was marked as a **track form**; for instance, the student might have decided to go *"fast in curves"*. The **speed description** code was assigned when the students' utterances contained words related to speed concepts; for example, the utterance *"fast, stop, fast, stop, four times"* would have received this code. The focus would be on the **speedometer** if the students described their plans using the speedometer's colours; for instance, the student might have said *"between fast and slow, not yellow"*. For the cases that are ambiguous or if the student does not give an utterance, the focus is set to **unknown**.

Observations: students' strategies

This part is a summary of Table 4.4 and Table 4.5. Although it was not clear from what kind of plans' modes students could have benefited from, the description of the plans' focus revealed an interesting strategy. In three types of focus (`speed description`, `speedometer` and `track form`) there was no guarantee that students were looking at the changes occurring on the graph. However, the `drawing graph` focus indicated that there was some coordination between how the students moved the car in order to observe certain output in the graph. For example, in order to make the graph *"go all the way up"* the student had to look at the graph to see if the plotted line really went to the top. In addition, "drawing" on the graph required the students to keep an eye on the track in order to maintain the control of the car. That is, a goal that required "drawing the graph" engaged the student in a behaviour where both representations need to be observed in

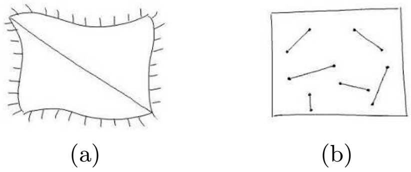


Figure 4.10: Pairs of cards from the pre-discrimination task.

order to produce the desired output. Therefore, this type of "drawing" interaction with the learning environment seems an interesting type of activity and it is investigated further in the next chapter.

4.5.4 Assessment method: students' knowledge of line graphs

Our second research question is regarding the assessment of students' interpretations. Our aim was to obtain more information of students' conceptions through their verbalisations in the Discrimination Task. However, only a few of the participants were willing to verbalise some of their thoughts during the task.

P21 was one of the students who thought aloud, and her utterances exemplify the type of verbalisations we observed. We noted that p21 was able to identify line graphs by name, but she might have applied her knowledge of line graphs to other items, too. For example, Figure 4.10a was sorted as a diagram while stating *"that's a line chart"*. Later, p21 sorted a picture as a diagram again (Figure 4.10b) stating that *"that one is a dot chart..."*. That is, p21 put attention to features she knew or recognised for the items she doubted and attempted to search for meaning in them. That is, it seems as if p21 was particularly ready to transfer her knowledge of line graphs to novel items.

It is possible that students did not widely express their opinions on the cards because there were too many cards to evaluate. This means that in an improved battery, students might need to spend more time with a smaller set of individual items in order to give them the opportunity to express their interpretations in more detail. That is, we will need to reduce the number of items the student examines/interprets.

Overall, although the assessment was useful to point out which type of learning environment might be more successful, it needs to be improved in a way that allows ut to obtain more students' verbalisations and gain more insight into students' knowledge as in the case above.

Importantly, we noted that students gave us richer interpretations of line graphs during the Select Track Questionnaire activity. Thus, their verbalisations were analysed in detail through a task analysis in order to get a better insight into how they interpreted the graphs.

4.5.5 Students' pictorial interpretations

Through the Select Track Questionnaire (STQ), we observed a rich variety of students' graphical interpretations. In particular, we came across what we called "mixed" types of interpretations. Before presenting our observations from the task analysis, we first show how students' verbalisations were coded and categorised.

Examples of students' utterances

Students' verbalisations were coded observing both students' utterances and students' track selection in the STQ. The coding scheme had five codes: `NO_UTTERANCE`, `AMBIGUOUS`, `NO_PICTORIAL`, `PICTORIAL` and `MIXED` (details are in Appendix C)

An utterance was coded as `PICTORIAL` if the abstract representation (the line graph) was confounded with a picture of a situation (the track). For example, these two utterances were marked as `PICTORIAL`:

- *"the track looks like the graph"*
- *"the graph is the same as the track".*

A `NO_PICTORIAL` code was given when the student interpreted part of the graph or all the graph correctly. That is, if the students assigned one or more relevant concepts (e.g., speed: fast; distance: kilometers) to the appropriate parts of the graph then it was said that students were not making a pictorial/iconic interpretation of the graph. For example:

- *"slower and then it's fast and then it's really fast"*
- *"it goes down, it has to be slower"*

A `MIXED` code was given if the students' verbalisations received a `NO_PICTORIAL` code and a `PICTORIAL` code simultaneously. For example:

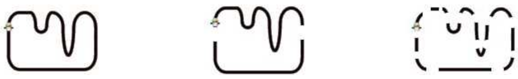
- *"This one because there's lots of up and down... probably the person that was doing it tried going really fast but they couldn't because they were in a wobbly track..."*

The first half of the sentence was coded as `PICTORIAL`, but the second part was coded as `NO_PICTORIAL` (see Appendix C).

A `NO_UTERANCE` code was given when the student did not say anything. In addition, if it was not possible to identify the referent the student was talking about (e.g., is the

(a) A graph. (b) One chunk. (c) Two chunks. (d) Several chunks.

Figure 4.11: A graph and some examples of how a student could have attended to its different segments/chunks.



(a) Track (b) (c)

Figure 4.12: A track and examples of how a student could have attended to its different parts.

student talking about the track or the graph?), the student did not provide any reason for his/her track selection (e.g., *"Probably that one"*), the utterance did not give an insight into how the graph is interpreted (e.g., *"I felt like"*), the student referred only to the track (e.g., *"it is a line"*), or the student only described the graph (e.g., *"It has lots of ups and downs"*) then the utterance was marked as `AMBIGUOUS`.

From this coding scheme, the relevant outcome came from the responses coded as `MIXED`, which can be treated as evidence of (1) how students' conceptions vary widely and (2) of the difficulty of assessing students' knowledge on the basis of their correct interpretations only. Thus, a task analysis was conducted in order to represent students' line graph interpretations in a way that we could observe possible forms of `MIXED` interpretations.

Task analysis of students' utterances

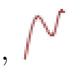
In summary, the task analysis (see Appendix D) describes an idealised process of a correct interpretation of a line graph (from the Select Track Questionnaire), but it also indicates in which parts of the process there is a possibility of giving a pictorial-like interpretation. From this analysis, we observed that students' utterances can be roughly described by the way students selected "segments" or "chunks" of the graph or of the track. Figure 4.11 and Figure 4.12 show some examples of how a graph and a track could have been chunked during the task.

The graph could be segmented in different ways in order to interpret it (e.g., Figure 4.11). However, some segmentations might be more effective for interpretation than others. For example, segmenting the plotted line as in Figure 4.11d might be more effective than segmenting it as in Figure 4.11b since each segment of Figure 4.11d could have an individual interpretation, but there is only one segment for Figure 4.11b and its interpretation using only one of the basic concepts (e.g., “fast”, “slow”, see Appendix D) does not seem adequate. Note that this does not mean that the segmentation as in Figure 4.11d is the only adequate way to segment/interpret the graph, but there might be some other configurations that would allow the student to interpret the graph correctly.

Thus, the way the student selects the chunks (“less” or “more” adequately) might affect the output of their interpretation. In particular, the segmentation of a graph could be interpreted more “globally” or more “locally” according to the “size” of the chunk. The semantics given to a larger chunk would seem as if the students were giving a general interpretation, whereas the semantics for a small chunk would look like a local interpretation of the graph. These ideas are further explained in the following two examples.

Example 1. For one of the interpretation questions, p03 seemed to select a less adequate segmentation of the graph and gave a global interpretation of it. For the graph in Figure 4.11a, p03 might have interpreted the graph choosing a configuration as in Figure 4.11c. P03 said when answering an interpretation question (a description of the gestures is underlined):

> This one {selects the track in Figure 4.12a}, because there’s lots of up and down up and down {hand moves up and down over the graph area, Figure 4.11a} and it probably the person that was doing it tried going really fast {index finger moves in zig zag - up and down} but they couldn’t because {index finger moves along the upper part of the selected track} they were in a wobbly track... so it was going slow down then really fast.

It seems as if p03 had inferred from the upper part of the graph (e.g., ) that a person tried to go “really fast”. That is, p03 interpreted part of the plotted line as “going fast”. Simultaneously, that graph segment was mapped to the upper part of the track (which looks like the graph); thus, p03 gave a pictorial interpretation as well. Overall, the utterance was coded as a `MIXED` interpretation.

Example 2. P35 seemed to have given a local interpretation for each of the segments of the graph. For example, p35 seemed to have chunked the graph in Figure 4.11a as in Figure 4.11d and interpreted it as follows:

> I think it went on that one {points to the track, Figure 4.12a}. Because it went up {hand moves away from the body} and as it went down it slowed

> down {hand moves close to the body}. And then as it went up fast {hand moves up, in front of the head}, it went faster and it went down [a bit] slower {hand moves down, in front of the body}. And as it went up it went faster again {hand moves up}.

The sequence of statements made by p35 seems to match "go up" with "fast" (e.g., the section of the graph that goes /); and "go down" with "slow down" (e.g., the section of the graph \ ). Thus, p35 gave local interpretations for each chunk of the graph, but simultaneously p35 selected the track that matches the plotted line and x-axis. This answer was coded as `MIXED` too.

Those two cases are examples of global and local interpretations mixed with pictorial answers. But note that there are many other possibilities of pictorial-like interpretations as the ones given by the two participants above. Overall, our observations from the task analysis and students' utterances point that students' interpretations seem to vary according to how the graph was chunked and how those chunks were interpreted.

This task analysis, however, is an idealisation of the interpretation process, it does not consider interpreting chunks in a different order (e.g., interpreting the graph starting from a compelling visual feature such as a sharp bend) and it imposes a strict order on the sequence of steps. Nevertheless, the task analysis is useful to comprehend the different cases where WYSIWYG p-prim could appear.

Observations from these qualitative analyses

A close analysis of students' utterances revealed new cases of pictorial interpretations, that is, of a wide range of instances when the WYSIWYG could be triggered. We noted that global and local interpretations could be obtained depending on which parts of the graph or the track, the students attended to.

These results might affect the way the assessment of students' line graph interpretations is performed. In particular, evaluation of students' knowledge might be more accurate when considering (these) alternative interpretations. In addition, we can note that the type of interpretation we observed was related to the assessment method we used; in other words, we might need to use different methods to evaluate students' knowledge. This approach will be tried in the next experiment.

4.6 Discussion and conclusion

Through this experiment we obtained valuable insights to answer our research questions. A discussion of this is presented.

4.6.1 Active vs. passive learning environments

Our aim was to observe whether an active learning environment might be more effective than a passive learning environment to improve students' knowledge of line graphs. The results of our analyses showed that students who used the active mode of the learning environment had a significant improvement in their scores from pre-test to post-test (see Section 4.5.2). Analyses also showed that students in the yoked passive modality did not have a significant improvement in their scores. Thus, this seems to suggest that participants who were able to manipulate the environment benefitted more from the intervention rather than those who watched their peers' experiments. Nonetheless, these results have to be considered with caution since we also noted that they might have been affected by a possible ceiling effect.

Note that observing that the active modality might be more beneficial than the passive modality is in agreement with other research. Beichner [9] said that in an embodied active mode, small event-graph units are created and could be managed in working memory as a single entity, thus, the *"simultaneous presentation of event and graph "makes the most" of the cognitive facilities available"* (p. 804). Our results agree with Anastopoulou, Sharples and Baber's [3] findings too. Anastopoulou et al. claimed that combining sensory modality and symbolic representations *"augment learners' experiences with the ability to physically manipulate symbols and see the effects of this manipulation as it occurs"* ([3], p. 267). They continued their argument stating that *"mapping changes of movement to variables... encourages learners' transition between concrete and abstract concepts"* ([3], p. 268). Thus, a possible explanation for our results is that if the student attended to the graph and track then through the manipulation of the car, several "units" were encoded (e.g., "fast:goes up", "slow: moves down", "stop: touches the bottom"). This, in terms of knowledge-in-pieces, might have augmented students' experiences and potentially helped to organise/connect students' knowledge structures in a more useful way – which in turn, might have encouraged transition to understand the graph abstractly.

Additionally, students in the active mode could have experimented with how their movements were reflected on the graph. Independently of whether the results met their expectations or not (e.g., the plotted line might have looked a way they did not expect), encoded "units" might have been added to their knowledge. Nonetheless, students in the passive mode had to make sense of the information presented by different sources (e.g., the graph and track). Therefore, they not only had to attend to different places, but they had to find out the relationship between the changes on the track and the changes on the graph – all this could potentially have increased the students' cognitive load (e.g., [65]).

"Drawing the graph" plan

We also investigated the students' plans within the interactive Racing Car activity in order to find out whether there were some particular activities from which students in the active modality might have benefitted from. No particular plans seemed to stand out, except the `drawing the graph` strategy. In this type of strategy, students raced the car but their aim was to observe "certain output on the graph" (e.g., *"two high peaks"*). Thus, this plan ensures that the student attends to the graph and track, whereas the other plans do not explicitly show that students were coupling the event with the graphical representation. In other words, the `drawing the graph` strategy might engage students to find out how the symbol is created by the physical movement. Therefore, we follow this lead in the next experiment.

Pictorial group

Note that an analysis of the pictorial group's performance did not show any significant improvement for any modality. However, we have to be cautious with this result since the number of participants in the pictorial group was small. Thus, in order to draw any conclusions, a larger group might be needed – this issue is addressed in the next experiment.

Additional observations regarding the learning environment

The interactive graphical learning environment was designed along the lines suggested by [60] and [70]. Thus, the Racing Car activity featured *dynalinking* – explicit interactive connections between abstract and pictorial representations, which highlight and facilitate comprehension. Although, in this experiment, no explicit measure was done on how dynalinking could have affected students' comprehension, there were indications that it could have been useful to pupils. On the one hand, some of the participants' utterances mentioned dynalinked features; thus it was used by students. On the other hand, it provided an alternative vocabulary to participants and it allowed them to express their concepts. For instance, some of them used the speedometer colours to express speed concepts (e.g., "yellow" as replacement for "slow"). Hence, this feature will not be removed for the next experiment.

4.6.2 Students' interpretations of line graphs

We tried to develop an alternative method to assess students' knowledge of line graphs. Although the assessment was useful to point out which intervention seemed more successful, it was not possible to use it to observe, in detail, students' conceptions since only a few of them produced enough verbal interpretations (perhaps, if this type of assessment is carried out, a reduced number of items would allow us to observe more verbalisations). During the Select Track Questionnaire (STQ), however, we observed rich explanations of how the students interpreted the graphs. Thus, we used the STQ to conduct a task analysis to explain and give insight into other possible states (or variations) of the students' line graph interpretations.

From local to global interpretations

The task analysis conducted revealed different points where a pictorial interpretation could occur. In particular, it indicated that segments or chunks of the representation could have simultaneous interpretations, both correct and pictorial-like. This suggests that depending on how the representation is chunked, there could be a continuum of pictorial interpretations ranging from "local" to a "global". Note that differences are treated as if they were within a "continuum" since the graph's segments could be very small in relation to the representation or they could be as large as the graph.

Our results describe students' pictorial interpretations more broadly than Clement's [16], but do not confront his framework. He considered only two "types" of the "graph-as-picture misconception" as global or as local errors: the full graph interpreted pictorially or a visual feature interpreted locally. Our analysis presents not only (1) a larger range of "local" and "global" interpretations, but also indicates (2) "double reading" of the graph's features (which could be simultaneously pictorial and correct).

Pictorial interpretations

So far, research literature does not clearly consider the wide range of varieties of pictorial interpretations. Our analyses of students' verbalisations on the STQ gave some evidence of `MIXED` interpretations. Particularly, the following variations were observed:

- "Pure pictorial" conceptions. For example:
  - "Plotted line as a picture": the plotted line is interpreted as a picture of an object.
  - "Plotted line and axis as picture": the plotted line and the x-axis are both interpreted as a picture of an object.
- "Mixed" conceptions: a correct interpretation and a pure pictorial conception happen simultaneously.
- "Local" to "global" pictorial interpretations.

Thus, our findings regarding uneven degrees of pictorial interpretations suggest that the status of the "graph-as-picture misconception" as traditionally conceived is inaccurate. Knowledge and misconceptions are fragmented ("in pieces", [64]) and the different forms of what is known as the "graph-as-picture misconception" can be considered as knowledge in transition rather than "erroneous" knowledge that should be replaced. Those fragmented pieces are not elicited equally by all forms of representation or are evident in all task contexts. They may also be transient and come and go depending on the cognitive load that a task imposes on the learner.

Overall, the interpretations given by students on the STQ gave some evidence of what can be considered as transitional states of students' knowledge. Evidence of these variations imply that the Graph Comprehension Questionnaire might have missed cases of other students' pictorial interpretations, and this suggests that using a variety of assessment methods (e.g., using both Discrimination Task and STQ) might be more useful to assess students and obtain more insight into their knowledge. Therefore, we need to improve the method of assessment and this is carried out in experiment 2.

4.6.3 Next steps

Given the details above, in the next experiment we focus on (1) investigating the effect of the active modality for learning about line graphs, (2) investigating the "drawing the graph" strategy, and (3) improving the assessment of students' interpretations of line graphs. In addition, we aim to monitor how the students' (pictorial) conceptions change by modality in order to confirm the role of the active mode as a possible tool to improve students' line graph comprehension.

Chapter 5

Experiment 2

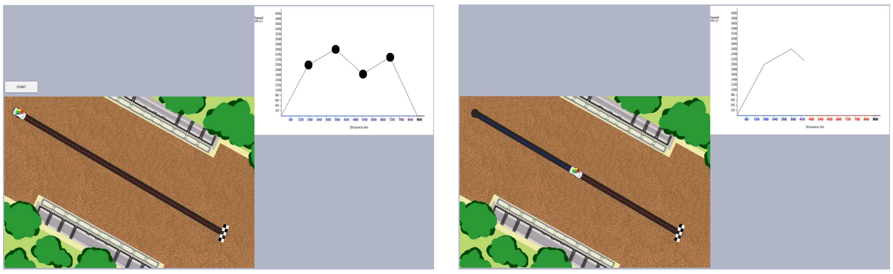
This chapter presents the design and application of an experiment that addresses two issues. Firstly, the results of the previous experiment seemed to support using the active environment in improving students' line graph interpretations. However, this result needed to be demonstrated again since it could have been due to a possible ceiling effect. Thus, in experiment 2, the active and passive task-interaction modes were administered to another group of students. In addition, an analysis performed in the previous experiment showed that the students used an interesting strategy during the Racing Car activity (the "drawing the graph" strategy); thus, we decided to investigate this by adding a new active mode (the *active_change_graph* condition) as well.

Secondly, after observing some of the students' pictorial interpretations, it was noted that a better assessment method was needed too. Therefore, we designed a new set of batteries to administer to students.

The results revealed that (1) the new *active_change_graph* condition holds promise as a learning environment that helps students to improve their line graph interpretations, and (2) a rich set of tests seems better to assess students' conceptions.

5.1 Introduction

Our two research questions were investigated further in this chapter: (1) Which form of interactivity is more effective for improving students' line graph interpretations? (2) How can we assess pupils' interpretations of line graphs? A summary of how we tackled each of them is in the next two subsections.



(a) Initial screen.

(b) Replay: the car is halfway through the track.

Figure 5.1: The active_change_graph condition of the Racing Car activity allowed students to handle the plotted line of the graph (left) and to race the car according to the students' graph configuration (right).

5.1.1 Active task-interaction modes

Results from experiment 1 suggested that, overall, students seemed to benefit more from the active task-interaction mode than from the passive task-interaction mode. However, an analysis on the performance of the pictorial group did not show the same result - possibly due to the small number of participants in that group.

Two lines of investigation derived from the analyses of the Racing Car activity nonetheless. The first line reinvestigates the effect of those modalities in order to confirm the results from experiment 1. The second line explores an alternative active mode.

The idea of a new active mode emerged when we observed that some participants in the active mode (experiment 1) seemed engaged in "drawing" the graph (e.g., *"I tried to get two big lumps"*, see Section 4.5.3). Thus, we designed another condition of the Racing Car activity where: (1) students could attend to the graph more; (2) students could keep the capability of deciding how to race the car; and (3) pupils could draw/make experiments on the shape of the graph. The new condition (*active_change_graph*) allowed students to handle the plotted line of a speed/distance graph. An example of this modality is in Figure 5.1. Students were encouraged to modify the graph by dragging up or down the black circles of the plotted line (Figure 5.1a). After the student had set the desired configuration, the "Start" button was pressed. Immediately afterwards, the system raced the car at the speed set by the student. Figure 5.1b shows a screenshot (taken half way through the replay) that corresponds to the configuration set in Figure 5.1a. In total, the race was shown twice; the aim was to make the active_change_graph condition comparable to the active mode from experiment 1 (in experiment 1, students raced the car once and then watched a replay of what they had done afterwards too).

Similar to the Racing Car activity in experiment 1, the active_change_graph condition was *dynalinked* (see Section 2.5.4 and Section 2.5.5). The *dynalinking* took the form of: (1) colouring the track and x-axis of the graph simultaneously as the car moved along the track (see Figure 5.1b); (2) the speedometer varied according to the car's speed; and (3) a background sound increased volume when the car went faster.

In addition, a passive modality of the active_change_graph condition was created. This modality presented students with video replays of what students in the active_change_graph condition did. That is, a yoked design similar to the one used in experiment 1 was applied: students in the passive mode watched the experiments of students who participated in the active mode. Since outcomes from the last experiment still need to be confirmed, it was necessary to complement the active_change_graph condition with its corresponding passive mode.

More details of these learning environments are provided in Section 5.3.2.

5.1.2 Varieties of students' conceptions

Results from experiment 1 showed that students' interpretations were complex. Examples of "pure" pictorial accounts were observed when interpreting line graphs (e.g., *"that graph sort of looks a bit like the track"*), but other students' verbalisations had to be classified as "mixed" due to the switches of pupils' answers between correct and pictorial conceptions (e.g., *"He walked fast and then he went down a bridge"*). In other words, the task analysis performed (Section 4.5.5) showed how rich students' knowledge could be and it highlighted some of the points where alternative pictorial/iconic interpretations could have happened.

In particular, we noted that there were "degrees" of pictorial conceptions since some students' interpretations were completely pictorial whereas others were partially pictorial. Additionally, some of those variations could have been considered "more" appropriate/acceptable than others. For instance, students who provided a mixed interpretation might have been closer to a more appropriate mental model for interpreting a line graph than students who answered with a pure pictorial conception. Recognising those differences could help to assess students' knowledge more accurately. As a consequence, we took another approach to assess students' interpretations during experiment 2, and we decided to create a diverse group of batteries in order to "capture" different students' conceptions. More details on this are in Section 5.3.1.

5.2 Hypotheses

In this experiment, we hypothesised that:

H1. Students who use any of the active modalities will interpret line graphs better than students who use any of the passive modalities.

We expected to confirm results from the previous experiment and to observe better performances for students going through the active modalities than students using the passive modalities. The passive modes required students not only to attend to both representations to discover their relationship, but to figure out if there were some particular aims that the driver wanted to achieve. Thus the passive modes could be less effective than their corresponding active modes.

We also hypothesised that (note that the active task-interaction mode from experiment 1 is referred to as the active_race_car condition here):

H2. Students' line graph interpretation in the active_race_car condition will differ from the students' line graph interpretation in the active_change_graph condition.

According to Beichner's account [9], the active_change_graph condition might not be as effective as the active_race_car condition. Beichner [9] advocates that *"real-time graphing lets the student process the event and its graph simultaneously rather than sequentially"* (p. 804). Thus, for the active_race_car condition, moving the car will create a unit between the pupils' hand motions and the changes in the graph; such units can be easily manipulated in the short-term memory [9]. However, under this theory, the active_change_graph condition does not create these series of motion-graph units, making it difficult for the student to attend to the changing events happening on the graph and track simultaneously; therefore, manipulating those events in the short-term memory would be less efficient.

5.3 Experimental design

Figure 5.2 presents an overview of the experimental design. The pre-test and post-test are formed by:

- The Graph Comprehension Questionnaire (GCQ).
- The Graph Interpretation Questionnaire (GIQ).
- The Graph and Tracks Questionnaire (GTQ).
- The students' graph interpretation test (Views).

The four conditions are referred to as: active_race_car[1], active_change_graph, passive_watch_car[2] and passive_watch_graph. In the passive_watch_car and passive_watch_graph conditions, pupils watched the experiments done by students who participated in the active_race_car and the active_change_graph conditions, respectively.

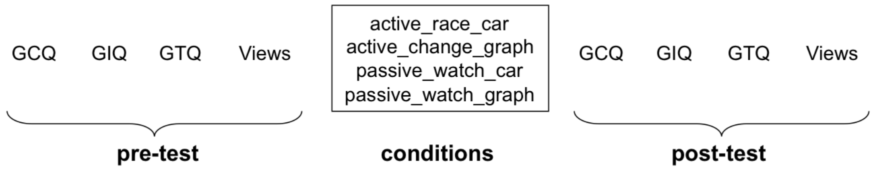


Figure 5.2: Overview of the experimental design. Graph interpretation questionnaires are part of the pre-test and post-test. The intervention consisted of one of four different conditions.

The pre-test and post-test contained batteries addressed to observe correct interpretations and variations of pictorial interpretations. The pre-test and post-test started with a "warming-up" assessment on general knowledge of line graphs (GCQ) and they continued with a series of interpretation questions (GIQ, GTQ, Views). The GCQ, GIQ and Views assessment tests differed in their format and all required the students to interpret different line graphs. The GIQ was a multiple-choice questionnaire with text-options as target items. The GTQ was a multiple-choice questionnaire too, but its target items were pictures of tracks (this battery was similar to the Select Track Questionnaire in experiment 1, see Section 4.3.2). The Views instrument was administered after the GIQ and the GTQ because it required students to verbalise their interpretations. Providing verbal interpretations could be a more challenging task for some students [37] and that is why it was left as the last assessment test.

Specific details of the pre-test, post-test and conditions are presented in the next sections. We start with the batteries' details for the pre-test and post-test and details of the conditions are presented afterwards.

5.3.1 Assessing students' interpretation of line graphs

We wanted to be able to measure students' line graph interpretations effectively, particularly for students who show the WYSIWYG p-prim (see Section 2.4.2). We had two tasks: (1) we needed to know when students' interpretations were correct and (2) we needed to take into account that students' interpretations could be a mixture of different pieces of knowledge (e.g., some of them could be correct, others pictorial, see Section 4.5.5) and we

[1]This condition was referred to as active task-interaction mode in experiment 1.
[2]This condition was referred to as passive task-interaction mode in experiment 1.

had to consider this during the assessment too. Thus, we decided to apply a different version of the discrimination task based on the outputs from experiment 1 (see Section 4.6.2) and we used additional assessment methods to observe some of the students' pictorial and mixed interpretations noted in experiment 1, too.

Thus, our target items included some of the pictorial variations we observed in the previous chapter. With this, we were able to observe, in more detail, how students' conceptions changed across conditions.

Before looking at each of the batteries, a summary of some of the pictorial interpretations that were observed is presented. The task analysis previously done (Section 4.5.5) was used as a tool for describing them and it was used to decide which pictorial interpretations to focus on.

Figure 5.3 presents an overview of a relevant section of the task analysis and it shows different alternatives for interpreting a line graph. An example about how to read Figure 5.3 follows. Imagine that somebody (Alex) is asked to interpret the graph that is on top of Figure 5.3. After looking at the relevant labels, Alex attends to the parts of the graph that will be used for the interpretation: for instance, Alex could focus on the plotted line (Figure 5.3, left side) or focus on the plotted line and the x-axis (Figure 5.3, right-side). Let's say Alex correctly selects the plotted line as the section to attend to. The next task consists of chunking the plotted line. If Alex chooses to chunk the plotted line in small segments (e.g., separating segments with positive slopes from segments with negative slopes) and assigns an appropriate interpretation to each chunk, then Alex is giving a correct interpretation of the graph (see *(1) specific interpretation*, Figure 5.3). However, if Alex assigns other interpretation to those chunks (e.g., identifying each segment as parts of a hill or a map) then Alex's interpretation could result in interpreting the graph as-a-picture of an object (see *(2) picture*, Figure 5.3) or as another representation such as a map (see *(5) map* from Figure 5.3).

Each path on Figure 5.3 corresponds to different interpretation answers, but only a few of them were explored further. The interpretations listed below were selected to cover some of the more common answers observed in experiment 1 - four of them were mixed/pictorial interpretations:

(1) **Correct:** a correct interpretation of a line graph (see *(1) specific interpretation*, Figure 5.3).

(2) **Pure pictorial:** a pure pictorial interpretation (see *(2) picture*, Figure 5.3).

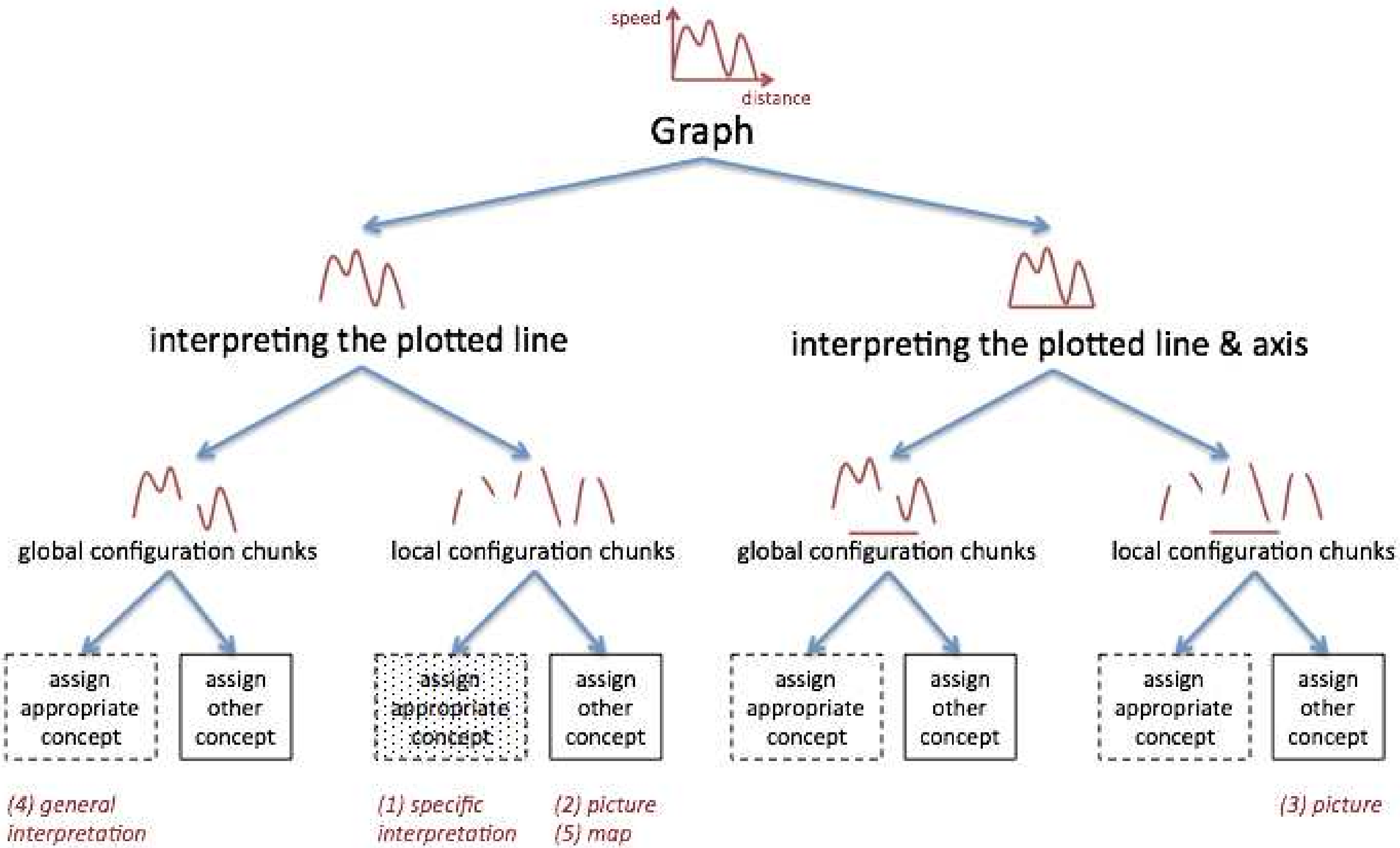


Figure 5.3: Interpreting a line graph: a representation of part of the task analysis performed in experiment 1.

(3) **Pure pictorial using x-axis:** a local interpretation where the x-axis is used as part of the plotted line (see *(3) picture*, Figure 5.3).

(4) **Mixed interpretation of the graph:** a mixed interpretation from the global features of the graph (Section 4.5.5, see *(4) general interpretation* in Figure 5.3). For instance, this  part could represent "racing fast" and "bumpy road" simultaneously.

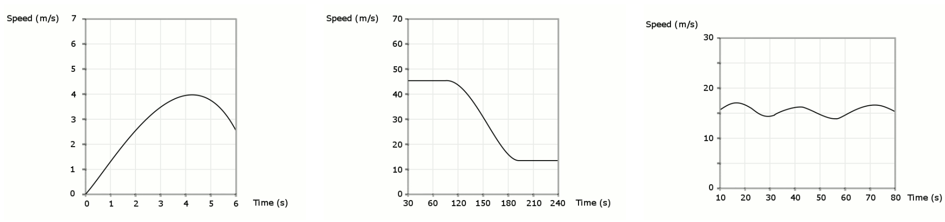
(5) **Graph-as-a-map:** An alternative interpretation that was not observed in the previous experiment, but other researchers have given examples of it, is the graph-as-a-map interpretation (see *(5) map*, Figure 5.3), for example, *"he went north and then south"*. Although the graph-as-a-map interpretation was not observed in experiment 1, it is included to account for the reported cases.

No more pictorial interpretations were selected because we wanted to avoid presenting students with a very large number of target items, particularly for the multiple-choice questionnaire which included text options. All or some of those five conceptions were used in each of the batteries that are presented next.

Graph Interpretation Questionnaire (GIQ)

A multiple-choice test (using text options as target items) helped us to observe correct interpretations and the existence of other pictorial interpretations explicitly. In other words, this method reduced ambiguity when pin pointing the student's interpretation. However, two of its drawbacks were that (1) specific pictorial interpretations had to be targeted and (2) the list of target items had to be short in order to keep the test manageable for young participants.

A total of three interpretation questions were presented to pupils and each interpretation question contained five different interpretation answers (the five interpretation forms that were discussed above). The three questions presented different line graphs: one of them showed a function with a predominant positive slope, another had a function with a predominant negative slope and the other had consecutive switches of slope (Figure 5.4). We used three questions for this test because we wanted to rule out the case where students selected their answer randomly. The probability of the students selecting the same type of correct/pictorial interpretation two or more times decreased depending on the number of questions presented and it gave more certainty of how the students were interpreting the graph.



(a) (b) (c)

Figure 5.4: Three different graphs used in the GIQ. Line graphs had (a) a predominant positive slope, (b) a predominant negative slope, and (c) repeated switches of slope.

The target items were created systematically. See Appendix F, Tables F.1, F.2 and F.3. In general, each sentence (or target item) in the pre-test had a parallel item in the post-test: the wording of each sentence was carefully matched between pre-test and post-test. For instance, the pure pictorial interpretation of the distracter in Figure 5.4c and of its parallel item were:

She cycled over a bumpy road.

She cycled over a wobbly road.

and the mixed interpretation for the item in Figure 5.4a and of its parallel item (which was a mirror image of Figure 5.4a, see Appendix F, Table F.2) were:

She cycled quickly to the top.

She cycled slowly to the bottom.

Details of this battery are in Appendix E, Table E.2.

Graph and Tracks Questionnaire (GTQ)

Another test with pictorial target items (pictures of tracks) was presented to students as well. Since the Select Tracks Questionnaire from experiment 1 permitted to observe a large variety of students' conceptions, it was decided to use it as part of the assessment this time.

It was not possible to tackle all four pictorial/mixed interpretations stated before (Section 5.3.1). Depicting the graph-as-a-map interpretation required a picture of a map, which could have stood out from the other target items (which were pictures of roads). Identifying a "global" or a "local" type of interpretation using pictorial items was problematic too and it was not clear how this could have been shown in pictures. Thus, the three conceptions left to represent were: a correct answer, a pure pictorial answer and a pure pictorial using x-axis answer. In order to create the target items, we translated possible text target items to their corresponding images. For example, a pure pictorial interpretation could have been read as *"it went up the road and then it went down"* and it could have been translated into a picture of a road going up and down. An example of one of the graphs and its corresponding target items is in Figure 5.5. Students were told that a driver drove on one of those tracks and were asked to select the track which corresponds to the graph. Option A in Figure 5.5 represents a pure pictorial answer. Option B represents a pure pictorial using x-axis answer. The correct option is C because the track has no bends and the car could have started slowly, speeded up and slowed down at the end. A fourth target item (option D) was added in order to increase and vary the target items presented.

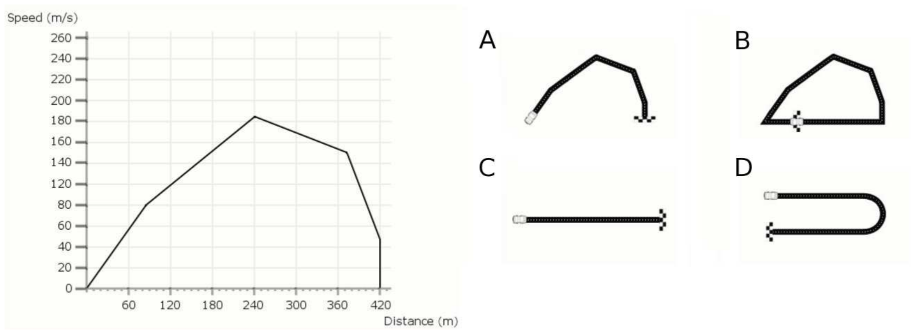


Figure 5.5: Sample of one of the questions and target items of the Graph Tracks Questionnaire.

Option D in Figure 5.5 presents the track that corresponds to the graph when the opposite interpretation of the graph is considered. For instance, Figure 5.5 shows a speed-distance graph, which could be interpreted as *"the driver goes slowly, then he speeds up and slows down at the end"*. The more appropriate track to choose, if the car is driven smoothly, would be the straight track: the car speeds up as it starts and slows down when it reaches the end. An inverse-incorrect interpretation would be *"the driver is going fast, slows down in the middle and then goes fast again"*. Considering that the car could go fast in straight segments and slow on bends, then the track to select for that inverse-incorrect interpretation is option D in Figure 5.5.

Note that the tracks show a "car" at the start of the track and a "finish line" at the end of the track. This design was derived from pilot tests (previous representations of the tracks, which did not include these details, were more difficult to consider as tracks).

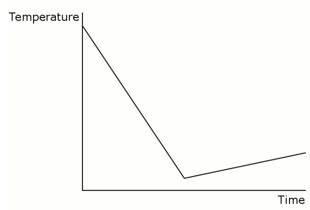
Similarly to GIQ, this test presented students with three interpretation questions, each one showed different graphs. Instructions and questions for this battery are in Appendix E, Table E.3.

In addition, note that the risk of using the Select Track Questionnaire (see experiment 1, Section 4.3.2) as assessment battery in this experiment was that the student could have some difficulties understanding the instructions, which asked the pupil to imagine a car racing on a track *before* being exposed to any of the conditions.

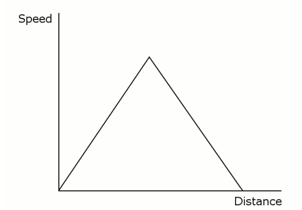
Students' graph interpretation test (Views)

The Views battery was used to obtain verbal interpretations from the students and it was based on observations from the Discrimination Task in experiment 1. It was noted that the Discrimination Task could have been more successful if more students had produced more utterances. Although providing them with interpretations could be more challenging to students than using production questions [37], requesting students' interpretations on a reduced number of line graph items could be more insightful of their conceptions and knowledge (e.g., [24]). Thus, this battery allowed exploration of other pieces of knowledge that may have not been observed before as well.

This test consisted of two items only: one item showed a speed-distance line graph and the other item showed a temperature-time or a noise-time line graph. That is, one of the items was within the context of kinematics and the other was in a different context (see Figure 5.6). Noise and temperature were chosen as alternative contexts because students were likely to be familiar with them: those concepts are everyday concepts and students might use these terms in school activities as well.



(a)



(b)

Figure 5.6: Each Views test had two items. Students had to interpret two different line graphs in two different contexts.

In particular, the items had two characteristics. First, one of the items showed a decreasing-increasing function (Figure 5.6a) and the other an increasing-decreasing function (Figure 5.6b). Second, the plotted line of one of the items was visually symmetric (Figure 5.6b), whereas the plotted line of the other item was not (Figure 5.6a). The same rules were applied to the items in the post-test battery; that is, if the speed-distance graph was symmetric and it showed an increasing-decreasing function in the pre-test then, in the post-test, it showed a decreasing-increasing function and the graph was not symmetrical.

This design was chosen in order to vary the presentation of the items within the test: (1) the graphs required "opposite" interpretations and (2) they showed different symmetries.

Each item appeared in the middle of the screen and students were asked for their interpretations by posing only one question: *What can you tell me about this?* Pupils were encouraged to explain as much as possible. After the students had interpreted the first item, a new screen with the second line graph was presented. The batteries for the pre-test and post-test and instructions are in Appendix E, Table E.4.

Graph Comprehension Questionnaire (GCQ)

The Graph Comprehension Questionnaire (GCQ) administered in the previous experiment was administered again. This time it was not only used to create our groups, but it was used as part of the assessment too since it could give us an initial measure of the students' general knowledge of graphs.

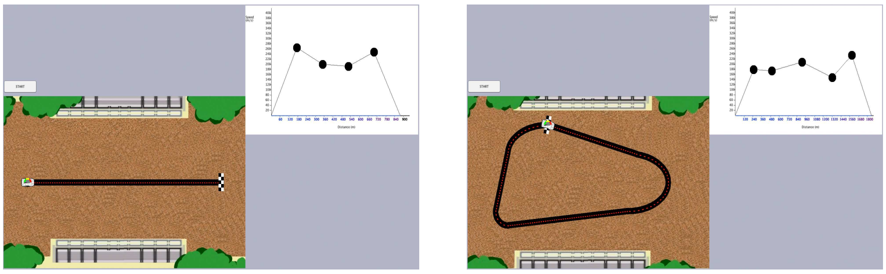
This test had the same characteristics as the Graph Comprehension Questionnaire from experiment 1, but with a slight modification. The test did not ask pupils to reflect on the graph before interpreting it. The Graph Comprehension Questionnaire in experiment 1 showed a screen (before the interpretation question) asking the students to reflect on the graph before continuing, but a diversity of behaviours were observed (some participants ignored the question while others tried to memorise the graph) and it was not clear if students took advantage of that step. The GCQ for this experiment and its instructions are in Appendix E, Table E.1.

Additional characteristics of the pre-test and post-test

While administering the GCQ, GIQ and GTQ, participants were not asked any questions in order to avoid them reflecting on their own answers.

Batteries were content-balanced too. For instance, a set of tests A (GCQ$_A$, GIQ$_A$, GTQ$_A$, Views$_A$) were administered to half of the participants before the intervention with the set of tests B (GCQ$_B$, GIQ$_B$, GTQ$_B$, Views$_B$) administered afterwards. The other half of the participants had the set of tests B administered before the intervention and the set of tests A afterwards. Additionally, the target items in the GCQ, GIQ and GTQ were randomly arranged in order to compensate for potential biases (e.g., selecting the first option all the time).

Note that each battery used different line graphs too. The GCQ showed distance-time graphs, the GIQ presented speed-time graphs and the GTQ showed speed-distance graphs.



(a) Graph 1: track without bends.

(b) Graph 2: track with bends.

Figure 5.7: Distribution of plotted line points for two graphs that correspond to two different tracks.

This should not hinder our objectives because all the graphs share the same context, and pictorial types of interpretations have been found in graphs that use those variables too (e.g., [10], [39], [7]).

5.3.2 Active vs. passive task-interaction modes

The two task-interaction modes of the Racing Car activity used in experiment 1 were administered again. For details of the Racing Car design see Section 4.3.2 and Section 3.1.2. Note that from now on, the active mode from experiment 1 will be known as the *active_race_car* condition, and its passive mode will be called the *passive_watch_car* condition. A description of the two new conditions is presented next.

Active_change_graph conditions

The aim of this condition was to help students to focus more on the graph and, at the same time, to keep the activity comparable to the active_race_car condition. Thus, the new active Racing Car activity had to allow students to freely manipulate the car through the graph without asking them to achieve any particular goals. Additionally, the axes, labels and tracks had to remain intact to maintain the same layout of the other conditions.

Mainly, the part we wished the students to interact with was the plotted line. We did not aim to convert the task into a "matching" task (e.g., "Graphs and Tracks", Section 2.5.5, Figure 2.5a), but to maintain a free manipulation of the plotted line (e.g., SimCalc MathWorlds®, Section 2.5.5, Figure 2.5b). Thus, it was decided to allow the student to modify the plotted line only, which was randomly generated and presented on the graph (see Figure 5.7). The student was able to freely manipulate the points of the line upwards or downwards, but not to the sides. After the pupils had selected the configuration of the graph, they were able to observe the result of their design by pressing the button "Start". Note that each of the black circles that gave shape to the plotted line was initially drawn around the middle part of the y-axis. In addition, the position of the points was either (1) evenly distributed along the x-axis (Figure 5.7a); or (2) they were distributed over the x-axis in an arrangement that allowed the student to race the car slowly on bends (Figure 5.7b).

When the car finished moving at the specified speed, a pop-up window appeared indicating that a replay was going to be shown. After the replay, a new button appeared and, when pressed, the next track was presented.

The students who used the active_change_graph condition observed the same tracks as the students in the active_race_car condition. Therefore, pupils saw six trials before starting the proper activity: during three of those trials the experimenter showed how to use the graph and the student practised on the other three (the experimenter and the student alternated trials). After those six trials were done, the students started the proper activity where they worked with six different tracks which varied in difficulty. See Table 5.1 and Table 5.2 for examples of the tracks and how the activity was demonstrated to the students.

The graph was dynamically linked to the track too. The active_change_graph condition kept the same dynalinking used in the active_race_car condition in order to make the two modalities comparable. Concurrent-verbal and retrospective-debriefing protocols were used too, in order to get an insight into the participants' strategies.

Note that Graphs and Tracks and SimCalc MathWorlds® have similar design characteristics to the ones planned for the active_change_graph condition (see Section 2.5.5). Both pieces of software (1) manipulate an animation of an event of the "real world" through an abstract representation, (2) are designed to enhance the learning experience, and (3) SimCalc MathWorlds® links different representations allowing the learner to potentially decrease the cognitive load (e.g., [60], [63], [40]). Positive results on the effectivity of those two learning environments indicate that the active_change_graph condition has the potential to facilitate students' learning as well.

Passive mode of the active_change_graph condition

Additionally, the passive mode of the active_change_graph condition (passive_watch_group condition) was implemented. Similar to the previous experiment, students' experiments during the active_change_graph condition were recorded in order to replay them to their peers during the passive_watch_graph condition. It was important to include the passive mode of the active_change_graph condition since we did not get a clear result about which

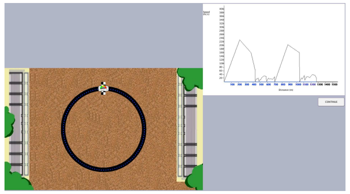
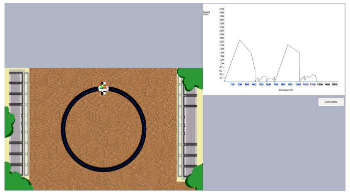
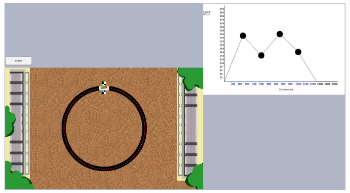
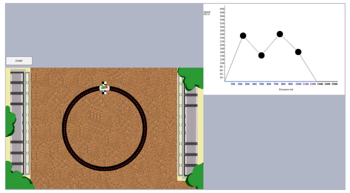
Active_race_car condition	Passive_watch_car condition	Active_change_graph condition	Passive_watch_graph condition
			
The experimenter raced the car and said:	The experimenter said, while the first replay happened:	The experimenter said, while changing the graph (the experimenter kept the black circles around the centre of the y-axis):	The experimenter said, when the replay happened:
"I am going to try to go fast sometimes and slow at other times, fast and slow."	*"I think this driver is trying to go fast sometimes and slow at other times, fast and slow."*	*"I am going to try to go fast sometimes and slow at other times, fast and slow."*	*"I think this driver is trying to go fast sometimes and slow at other times, fast and slow."*

Table 5.1: Examples of how the experimenter demonstrated how to use the learning environment to the students during one of the trials for the four conditions.

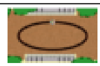
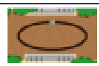
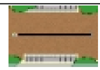
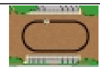
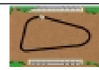
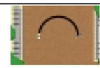
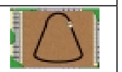
	Trials						Activities					
Tracks:												
User:	E	S	E	S	E	S	S	S	S	S	S	S

Table 5.2: Sequence of tracks presented during the trial activities and the proper activities during the Racing Car activity. E and S stand for experimenter and student, respectively.

modality was effective, thus we had to complement the active mode with its corresponding passive modality. Instructions for this activity are in Appendix E, see Figure E.3 and Figure E.2.

5.4 Method

5.4.1 Participants

A total of 38 primary school students in the fifth school year participated in the activities. Their age ranged between 9;5 years to 10;4 years old (M = 9;11 years) and 17 of them were male and 21 of them were female. The Deputy Head of the school selected the students from five different classrooms and from different ability groups – children who the teachers thought might exhibit problem behaviour did not participate in the study. Thus, the participants could be considered to be a representative sample of students with different abilities.

5.4.2 Procedure

The experiment took place in a (state) primary school located in a middle-class area. A small room next to one of the classrooms was used to interview the students. A little noise from the class next door could be heard, but it was sufficiently quiet. Teachers came into the room at times, but the interruptions were minimum. The infra-red finger tracking (IRFT) system was horizontally positioned over a table with the Wii remote control hanging about a metre over the screen (see Section 3.2.2).

The students were seen individually. For each session, the parental consent form was checked and, if any type of recording was permitted by the parent, the student was asked for his/her consent as well. A brief explanation of the tasks was given and the participant was reassured that the activities could be stopped at any time. Afterwards, the infra-red light of the IRFT system was positioned on the student's dominant hand and the experimenter showed him/her how to use it. A few minutes were spent drawing using a painting application in order to get the participant familiarised with the system. If necessary, a calibration of the IRFT system was done using the Wiimote software.

After the drawing task, the activities were presented to the students in the order stated above (Section 5.3). At the start of each activity, the students were asked if they were happy to read the instructions by themselves or if they wanted the experimenter to read them. During the task, no feedback of the activities was given to the participants, but they were always encouraged to answer as best as they could. At the end of the activities, the students were thanked and asked to bring the next participant to the room.

5.4.3 Independent and dependent variables

Independent variables

There were two between-subjects factors: *task-interaction mode* (or *mode*) and *task type*. Each one had two levels:

- The *task-interaction mode* had two levels: active and passive.
- The *task type* had two levels: car_material and graph_material.

Thus, there were four conditions: active_race_car (active mode and car_material task type), active_change_graph (active mode and graph_material task type), passive_watch_car (passive mode and car_material task type), and passive_watch_graph (passive mode and graph_material task type). There was only one within-subject factor: *time*.

- For *time*, we had two levels: pre-test and post-test.

Dependent variables

The dependent variable was the aggregated scores obtained from the GCQ, the GIQ and the Views battery (see Section 5.5.2).

5.4.4 Distribution of participants across conditions

Students were assigned to one of four different conditions (groups): active_race_car, active_change_graph, passive_watch_car or passive_watch_graph. Care was taken to keep the groups as balanced as possible with respect to the number of males and females. Students from different ability groups were placed in each of the four groups as well (we were given a rating of each pupils' abilities and we tried to make sure there were equal numbers of each ability level in each group).

5.5 Results

Data from five of the 38 students who participated had to be excluded from the analyses. Three of those five students needed some help to read and/or to interpret at least one of the line graphs presented during the pre-test. Other students needed help interpreting at least one of the graphs and did not complete the post-test. Another student declined to answer the interpretation questions for the GIQ and the GTQ during the pre-test and the post-test. It was considered that the feedback given by the experimenter could have affected those students performances and, for that reason, those cases were not considered during the analyses.

5.5.1 Groups of participants

The following analyses were undertaken using the data from 33 participants (19 female and 14 male). The number of participants in each condition is presented in Table 5.3.

Also note that, as in the previous experiment, we will differentiate the two groups: a *pictorial group* and a *non pictorial group*. These groups were formed by the way students chose their answer to the first interpretation question that was posed to them (which was from the GCQ, see Section 5.3). Thus, the pictorial group consisted of students who chose any pictorial answer in the GCQ, the rest of the students were in the non pictorial group. This is not one of the independent variables, but it is a way to observe the effect of the modalities in a subgroup of our participants (specifically, the pictorial group).

In order to find out if the students in the pictorial group selected their answers by chance, a goodness-of-fit χ^2 test was performed. Results showed that the number of students selecting any of the pictorial answers in the GCQ differed significantly from the expected, $\chi^2 = 9.78$, $df = 1$, $p < 0.01$. This shows that the pictorial group favourably represented the students who seemed to answer with a pictorial interpretation predominantly.

Condition and Mode (all students)	N	Total
active_race_car	11	33
active_change_graph	9	
passive_watch_car	7	
passive_watch_graph	6	
active	20	33
passive	13	

Condition and Mode (pictorial group only)	N	Total
active_race_car	7	22
active_change_graph	5	
passive_watch_car	5	
passive_watch_graph	5	
active	12	22
passive	10	

Table 5.3: Number of participants for each group.

In the following sections, we start by describing the scoring for each of the tasks (Section 5.5.2). Results regarding the first research question are in Section 5.5.3 and analyses for our second question are in Section 5.5.4.

5.5.2 Scoring

Each of the batteries was first scored individually before obtaining an aggregated score. We start presenting how we scored each of the batteries and end with details of how the aggregated scores were obtained.

Scoring the "Graph Comprehension Questionnaire"

A point was given for a correct interpretation answer otherwise no points were given. The maximum score for a student was 1 and the minimum was zero.

Scoring the "Graph Interpretation Questionnaire"

For each of the interpretation questions one point was given if the student answered correctly. The maximum score for a student was 3 and the minimum was zero.

Scoring the "Graph and Tracks Questionnaire"

Results of this test had to be eliminated from the analyses. The majority of the participants selected either the track that resembled the plotted line of the graph, or the track that resembled the plotted line with the x-axis of the graph. This behaviour was observed for the pre-test and the post-test of the Graph and Tracks Questionnaire. Only three students answered one of the questions of the pre-test correctly and one student answered one question of the post-test correctly.

The overwhelming pictorial responses made clear how fragile the students' knowledge was. In spite of students exercising how to interpret other line graphs in the GCQ and GIQ batteries (Section 5.3), participants seemed highly distracted by the similarity of the target items with the graph (even after the Racing Car activity had taken place!). These results contrast highly with the rich verbalisations and variety of responses observed for a similar battery administered in experiment 1 (Section 4.5.5). The effect of presenting this test before the intervention did not prove fruitful and it shows that this task might be more suitable as part of the intervention tasks (as Janvier originally proposed [38]).

Scoring the "Views" test

Coding In order to assess students' utterances, a coding scheme was first developed. It took some effort to develop a simple and coherent coding scheme. However, in order to describe some of the main characteristics of the students' conceptions in a comprehensive way, it was decided to use three coding dimensions. Each coding dimension consisted of a binary code and each student's interpretation received one and only one code from each coding dimension. The coding dimensions were independent of each other; thus, codings could be analysed either separately or in combination, depending on the research question. See Figure 5.8.

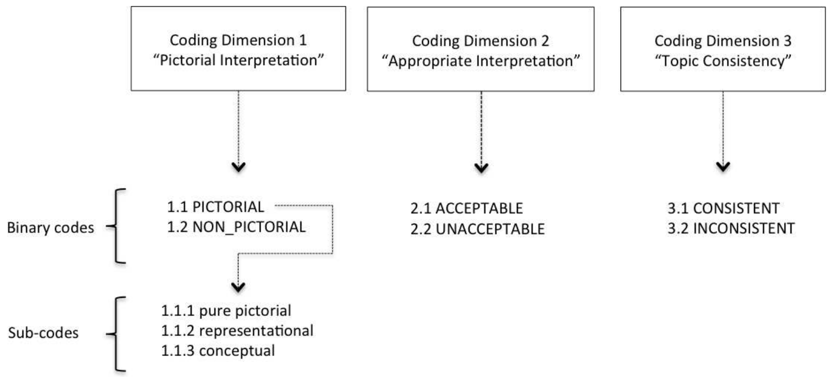


Figure 5.8: An overview of the coding scheme and its three dimensions.

Coding Dimension 1 screened students' pictorial conceptions and it gave a basis to observe other pictorial interpretations. It had two codes: `PICTORIAL` and `NON_PICTORIAL`. For each utterance coded as `PICTORIAL` a sub-code was given too. There were three sub-codes: `pure pictorial`, `representational` and `conceptual`. Briefly, the `pure pictorial` sub-code was given to an interpretation that confounds the plotted line as a real world or abstract object (e.g., a hill, a triangle); the `representational` sub-code was assigned when the graph was interpreted as another external representation (e.g., a map); and the `conceptual` sub-code was given when the graph was interpreted as a picture of an abstract concept (e.g., a picture of time). More examples of each sub-code are in Appendix G.2.

In order to provide a more comprehensive view of the various forms of pictorial interpretations, the second coding dimension was used to mark if the student verbalised at least a single appropriate interpretation of the graph. A correct interpretation using either both or one of the variables of the graph was considered appropriate.[3] The codes used were `ACCEPTABLE` and `UNACCEPTABLE`. This coding dimension worked in combination with Coding Dimension 1. By using both dimensions we captured evidence of "mixed" interpretations; that is, it allowed observing utterances that had an `ACCEPTABLE` code and a `PICTORIAL` code at the same time. Details of this code are in Appendix G.3.

Not all the students' utterances were consistent. Some students interpreted the graph in another context to the one provided. In order to give an account of those cases, a third

[3]Students who struggled interpreting one of the variables in terms of the other, but were successful when interpreting them separately received an `ACCEPTABLE` code. In other words, as soon as one of the variables was interpreted correctly then the `ACCEPTABLE` code was used.

coding dimension was used to mark CONSISTENT and INCONSISTENT utterances. A more detailed description of this code is in Appendix G.4. This coding dimension was used to analyse the students' responses without inconsistent cases.

In general, the three different coding dimensions provided an in-depth study of the different types of pictorial interpretations. Details of each coding dimensions are in Appendix G.

Note that, as specified in each coding dimension, a NON_PICTORIAL, UNACCEPTABLE or CONSISTENT code did not guarantee that the student had given a useful utterance (e.g., What can you tell me about this? *Nothing*). The rest of the codes indicated that there was evidence of a PICTORIAL, an ACCEPTABLE, or an INCONSISTENT answer.

Scoring Each of the students' utterances were coded using the coding scheme described above. A student could get a maximum of two equal codes (e.g., two PICTORIAL codes) and a minimum of zero codes for each coding dimension (e.g., zero INCONSISTENT codes).

A point was only given if the student got an ACCEPTABLE code *and* a NON_PICTORIAL code simultaneously – that is, a point was given if there was evidence of a correct interpretation, which was not pictorial. Thus, the maximum score for this test was two and the minimum was zero.

Aggregated scores

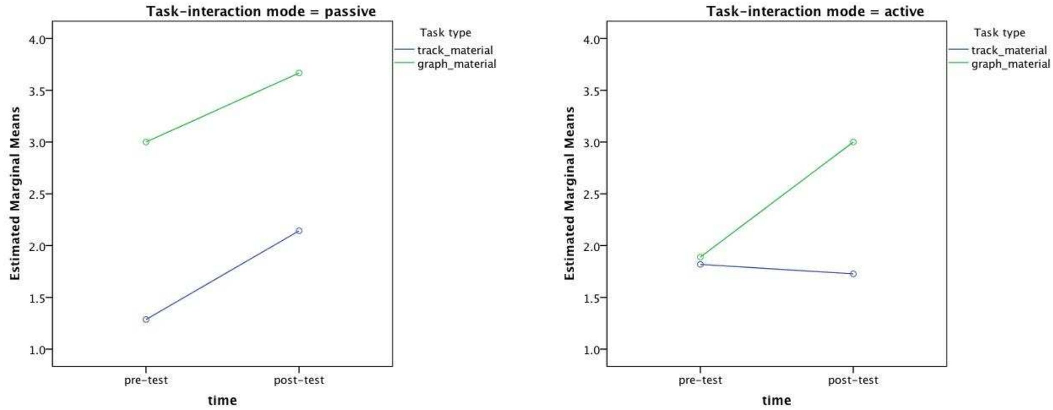
Scores from the GCQ, GIQ and Views were added up. Scores could have been between zero and one for the GCQ; zero to three for the GIQ; and zero to two for the Views battery. Therefore, the aggregated scores ranged between zero to six.

5.5.3 Active vs. passive modes of the learning environment

We wanted to compare which modality was more effective for improving students' line graph interpretations (using the aggregated scores). We checked this by looking at the performance of all the pupils, but we analysed the results for the pictorial group too.

Active vs. passive modes: all students

We conducted a 2×2×2 three-way mixed ANOVA (task-interaction mode [active, passive] × task type [car_material, graph_material] × time [pre-test, post-test]) on students' aggregated scores. There were not any significant interaction effects. However, we found a significant main effect for time: Wilks' Lambda = 0.74, $F(1, 29) = 10.2$, $p = 0.003$,



(a) Passive mode.

(b) Active mode.

Figure 5.9: Aggregated scores by task type and by task-interaction mode.

multivariate partial $\eta^2 = 0.260$; and there was a significant main effect for task type: $F(1{,}29) = 5.39$, $p = 0.027$, partial $\eta^2 = 0.16$. There was no significant main effect for task-interaction mode. Figure 5.9a and Figure 5.9b show the means for the car_material and graph_material task types by task-interaction mode.

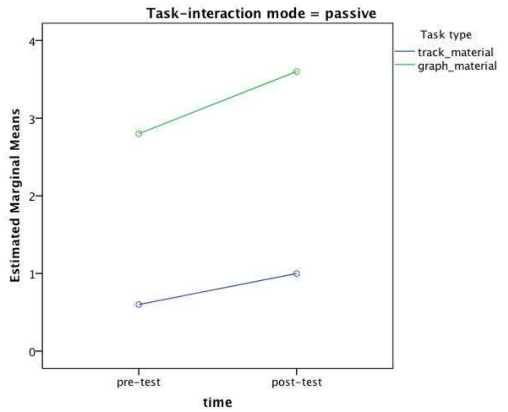
Paired sample t-tests were performed on the aggregated scores for the four conditions to look in more detail at which condition seemed to contribute to the observed significant main effects. There were no significant effects for the passive_watch_car, passive_watch_graph and active_race_car groups. However, the active_change_graph condition (active mode & graph_material task type) had a significant improvement from pre-test to post-test [$t(8) = -3.16$, $p = 0.013$, $\eta^2 = 0.56$].

Therefore, only the students who went through the active_change_graph condition seemed to improve their interpretations of line graphs.

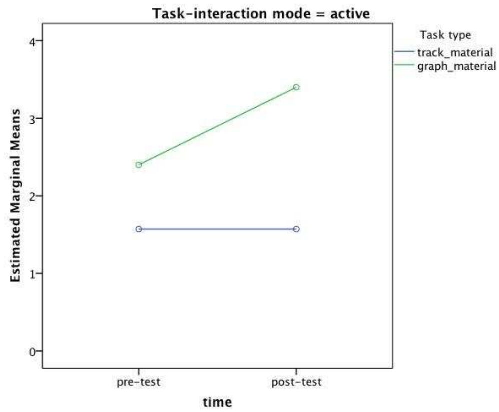
Active vs. passive modes: pictorial group only

A 2×2×2 three-way mixed ANOVA (task-interaction[active, passive] × task type [car_material, graph_material] × time [pre-test, post-test]) was carried out on the aggregated scores for the pictorial group only (see Section 5.5.1). There was a main effect for time: Wilks' Lambda = 0.772, $F(1{,}18) = 5.33$, $p = 0.033$, multivariate partial $\eta^2 = 0.23$. There was a significant main effect for task type too: $F(1{,}18) = 12.47$, $p = 0.002$, partial $\eta^2 = 0.41$. There were not any other significant main or interaction effects. See Figure 5.10a and Figure 5.10b.

Paired sample t-tests were performed on the four conditions to look in more detail at which group seemed to contribute to the significant main effects that were observed. There were no significant effects for the passive_watch_car, passive_watch_graph



(a) Passive mode.



(b) Active mode.

Figure 5.10: Aggregated scores by task type and by task-interaction mode: pictorial group only.

and active_race_car groups. However, the active_change_graph condition (active mode & graph_material task type) had a significant improvement from pre-test to post-test [$t(4)$ = -3.16, p = 0.034, η^2 = 0.71]. Hence, students in the pictorial group only benefited from the active_change_graph condition.

Results overview

Importantly, note that there was no interaction effect between the task-interaction mode and time, and there was no main effect for the task-interaction mode too (neither for the whole group of students nor for the pictorial group). This suggests that there is no significant difference in students' performance between active and passive modes.

However, there was a main effect for task-type and for time. Further analyses revealed that, from the four conditions (mode $\times$ task type), the active_change_graph condition was the only one which showed a significant increment of correct scores (i.e. there was a decrease of pictorial interpretations). Furthermore, the active_change_graph condition was useful for the pictorial_group too. Interestingly, the students who went through the active_race_car, passive_watch_car and passive_watch_graph conditions did not show a significant improvement in their line graph interpretations.

5.5.4 Assessing students' line graph interpretations

We used different tests for assessing students' interpretations and each one seemed to give us a different view of students' knowledge. That is, when a chi-squared test was performed to explore whether there was an association of students' correct answers between any two batteries (e.g., did students who selected a correct option in the GCQ give a correct answer in the GIQ?), results did not show any significant association across any of the tests (GCQ×GIQ, GIQ×Views, Views×GCQ). This suggested that different students' conceptions were picked up by different tests. In other words, each of the batteries was sensitive enough to detect different aspects of the students' knowledge; therefore, by aggregating the scores across tests we obtained a richer measure of students' interpretations than using any of the batteries alone.

Interestingly, the different batteries gave us some information of how students' conceptions changed and of pictorial variations we had not considered before as well.

Which of the students' pictorial interpretations changed? Each battery allowed observing different answers from students, thus we had to group students' answers into categories that could be clearly identified across batteries. Responses were classified as: *correct*, *mixed* and *pure pictorial* interpretations. Any answer that did not fall on those categories was classified as *other* type of interpretation (e.g., the "other" target item from the GCQ).

For each student, we created a frequency table with the student's answers across batteries (for the pre-test and post-test). For example, Table 5.4 shows the type of answers participant P01 gave for each battery.

	pre-test				post-test			
	GCQ	GIQ	Views	Total	GCQ	GIQ	Views	Total
pure pictorial	1	2	0	3	1	3	0	4
mixed	0	1	0	1	0	0	0	0
correct	0	0	2	2	0	0	2	2
other	0	0	0	0	0	0	0	0

Table 5.4: Frequency of a participant's answers across batteries.

The students' answers were aggregated and the total amount from the pre-test was compared with the total amount of the post-test. This was done for: (1) all the students, (2) the pictorial group and (3) the students who underwent the active_change_graph condition; see Table 5.5. We also present the frequencies for the (4) active_race_car, (5) passive_watch_car and (6) passive_watch_graph conditions; see Table 5.6.

Consider Table 5.5, overall, the proportion of change was the greatest for the correct answers. It is also possible to appreciate that the frequency of correct answers increased from pre-test to post-test by 60% for the students who underwent the active_change_graph condition and by almost 30% for all the students and for the pictorial group. It seems from Table 5.5 that the intervention might have predominantly helped students giving mixed

	All students			Pictorial group			Active_change_graph group		
	pre-test	post-test	post-test ÷ pre-test	pre-test	post-test	post-test ÷ pre-test	pre-test	post-test	post-test ÷ pre-test
pure pictorial	65	67	1.03	50	50	1.00	17	16	0.94
mixed	47	32	0.68	32	20	0.63	15	8	0.53
correct	64	83	1.30	40	51	1.28	17	27	1.59
other	22	16	0.73	10	11	1.10	5	3	0.60

Table 5.5: Students' answers across three different (sub)groups.

	Active_race_car group			Passive_watch_car group			Passive_watch_graph group		
	pre-test	post-test	post-test ÷ pre-test	pre-test	post-test	post-test ÷ pre-test	pre-test	post-test	post-test ÷ pre-test
pure pictorial	28	32	1.14	13	16	1.23	7	3	0.43
mixed	13	10	0.77	11	5	0.45	8	9	1.13
correct	20	19	0.95	9	15	1.67	18	22	1.22
other	5	5	1.00	9	6	0.67	3	2	0.67

Table 5.6: Students' answers across three different groups.

interpretations to make a transition towards the correct interpretation. However, it is not possible to say the same for the cases where students gave a pure pictorial answer.

Consider Table 5.6, which shows the answers for those groups that did not show a significant improvement of students' correct interpretations in Section 5.5.3. Interestingly, the proportion of correct answers for the active_race_car group did not seem to increase; however, the corresponding passive mode of this group (i.e. the passive_watch_car condition) showed an increase in correct answers by 67%. In addition, note that the passive_watch_car group was the only one where there was a large decrease in the proportion of pictorial answers and "other" types of answers across all groups; this means that students' conceptions seemed to have changed to mixed or to correct interpretations.

Overall, we can see that the different conditions seem to differ in how they help students to understand line graphs. Importantly, (1) not all active modes seemed to have a positive effect on students' interpretations and (2) students who gave mixed interpretations might have benefitted by some conditions (e.g., active_change_graph condition).

Pictorial variations The Views battery revealed other types of interpretations that were not seen previously. As it was already described in the coding scheme, the Views battery could code mixed interpretations (marked as `PICTORIAL + ACCEPTABLE`) and pure pictorial interpretations (marked as `PICTORIAL + UNACCEPTABLE`). However, some of the pictorial interpretations turned out to be "conceptual". As stated in the coding scheme in Appendix G.2.3: the conceptual code was given when a pictorial representation of an abstract concept happens:

> For example, a student who points at the plotted line and interprets it as *"time goes up and down"*, shows that (1) the student is making explicit the object he/she is talking about: the line that comes "up and down"; and (2) the line is the representation of the abstract concept "time". Thus, the concept takes shape on the graph.

The proportion of students' interpretations can be observed in Table 5.7 and Table 5.8 (frequencies were based on the 66 items seen by each of the 33 participants in the pre-test). It can be observed from Table 5.7 that about 30% of the utterances were coded as PICTORIAL and about half of those interpretations were coded as MIXED. Table 5.8 presents the percentage observed for the PICTORIAL sub-codes. It shows that a relatively large proportion of the total pictorial interpretations were conceptual interpretations (about 1/3 of them). In spite of that, we could not observe this kind of interpretation through the other assessment batteries nor in the previous experiment. Thus, using a richer set of batteries could reveal students' conceptions that have not been considered before. This also supports that the traditional definition of the "graph-as-picture misconception" does not always reflect students' types of interpretations.

		Coding Dimension 2 (%)		
		UNACCEPTABLE	ACCEPTABLE	Total
Coding Dimension 1(%)	NON_PICTORIAL	13.6	56.1	69.7
	PICTORIAL	13.6	16.7	30.3
	Total	27.2	72.8	

Table 5.7: Proportion of MIXED and PICTORIAL conceptions.

		Coding Dimension 2 (%)		
		UNACCEPTABLE	ACCEPTABLE	Total
PICTORIAL (%)	pure pictorial	10.6	9.1	19.7
	representational	1.5	0.0	1.5
	conceptual	1.5	**7.6**	9.1
	Total	13.6	16.7	30.3

Table 5.8: Proportion of the types of PICTORIAL cases.

5.6 Discussion and conclusion

We had two objectives for experiment 2. The first objective was two fold: (a) to confirm whether the active task-interaction mode is an effective way to improve students' interpretations of line graphs; and (b) to test a new active mode (active_change_graph condition). Our second aim was to improve our assessment of students' line graph interpretations – particularly for students who interpreted the graph pictorially. A discussion of the results obtained is presented next.

5.6.1 Confirming results from experiment 1

In experiment 1, students seemed to have performed better under the active mode than under the passive mode and we hypothesised a similar outcome for experiment 2. However, analyses conducted on the active vs. passive task-interaction modes did not support results from experiment 1. In other words, we did not find evidence supporting that the active task-interaction mode is a more effective environment than the passive mode for improving the students' interpretations – neither for our whole group of students nor for the students in the pictorial group.

In spite of that result, we also observed that one of the active task-interaction modes seemed successful. Given this, we could infer that not *all* types of interactivity might be beneficial, but only *certain* types of interaction could be more useful for supporting line graph interpretation.

Thus, why did we observe some improvement in experiment 1? There were different factors that could have made its active mode more successful, e.g., participants' age range, the assessment task measure was different, the active mode in experiment 1 had two tasks. However, we have to remember that some participants in experiment 1 purposely decided to "draw" the graph. That is, the group in the active mode (experiment 1) included students who used the "drawing graph" strategy (4 of the 12 students in the active mode for experiment 1 showed evidence of "drawing the graph"), which could have contributed to the improved performance that was observed in experiment 1.

Overall, we cannot confirm our hypothesis (see Section 5.2) that all active modalities help students interpret line graphs more than the passive modes.

5.6.2 Improvement of performance for the pictorial group under the active_change_graph condition

Interestingly, analyses on the aggregated scores showed a general improvement of students' interpretations of line graphs. However, the improvement seemed to have come from the active_change_graph condition only. This was the case when considering results of (1) all students and of (2) the pictorial group only (in spite of the small number of students in the latter group).

Although there was not a control condition to compare against the active_change_graph condition, the improvement could not be due to a test re-test effect. If a test re-test effect had happened, we would have obtained a significant improvement across all the conditions; but this was not the case.

Thus, students who underwent the active_change_graph condition showed a significant improvement in their interpretations of line graphs, but this also means that they reduced their pictorial interpretations, particularly mixed types of pictorial interpretations. Overall, this result highlights the active_change_graph condition as a promising method for helping students to improve their line graph interpretations.

5.6.3 The active learning environments

There are a couple of points to discuss from these results. The first point is regarding the difference in students' performance between the two active conditions (active_race_car and active_change_graph). The second is regarding the role of the active_change_graph condition as a graph-construction task.

Dynamic generation of hypotheses vs. planned hypotheses

One of the interesting results is the difference in performance observed between the active_race_car and active_change_graph conditions. Both conditions encouraged students to freely play with the system. That is, students had some control on how they wanted to race the car. However, the active conditions differed on (1) the degree the students could change plans and (2) the parts of the learning environment that students had attended to.

Regarding the first point, both active conditions differed on how students developed their plans or hypotheses. In the active_change_graph condition participants had to make a decision or plan the car's movement *before* racing the car. The plan did not change once the race started. As a consequence, the learner could spend time observing how the graph's patterns were reflected on the cars' movements.

The active_race_car condition, however, allowed planning and testing hypotheses *before* or/and *during* the race. That is, there were more opportunities to test several on-the-go hypotheses. More hypotheses-testing in the active_race_car condition could imply more experiences to add to the students' knowledge experience and, in terms of the knowledge in pieces perspective [27, 64], this could have been beneficial. However, part of the students' performance may have relied on (1) how systematically those hypotheses were organised (or tested) and (2) how deeply the outputs were processed (particularly when several changes of plan occurred). These were some aspects that could have explained why the active_race_car condition was not as effective as the active_change_graph condition.

Regarding the second point, participants in the active_race_car condition could have decided whether to attend (a) to the "track" or (b) to the "graph and track" at any time of the activity. This meant that a student could have raced the car without looking at the corresponding changes in the graph; thus, reducing his/her opportunities to assign meaning to parts of the plotted line. In contrast, students who underwent the active_change_graph condition were "forced" to attend to the graph and to make a decision on the graph's configuration at the start of the activity. That is, independently of students' plans, they were more likely to engage in the meaning (either correct or incorrect) of the plotted line as they modified it. Students in the active_change_graph condition might or might not have looked at the graph again, but at the very least, all of them attended the graph once in every trial.

Embedded graph construction

Although the active_change_graph condition is not like a traditional teaching tool (i.e., there are no specific goals, no marks, no correct or incorrect answers), it seems to have played an instructional role. In some way, the task resembled a "construction" task; although students did not have a say on how labels, scales or axes were set, pupils could decide how the plotted line might look like. Furthermore, the active_change_graph condition resembled Ainley's [1] research (see Section 2.3.1) since it served as a tool of exploration too. Thus, this task allowed students to explore the graph and therefore, engaged students in an interpretation task. Agreeing with Leinhardt et. al [45]:

> both construction and interpretation can vary with respect to the features that are being attended to (local-global, quantitative-qualitative). In terms of their relationship to each other, it can be noted that whereas interpretation does not require any construction, construction often builds on some kind of

interpretation (p. 13).

5.6.4 Observations from students' line graph interpretations

Each of the tests allowed us to observe several students' interpretations. Besides noting the well-known "pure" graph-as-picture responses, we had the opportunity to investigate in more detail the students' mixed responses. Mixed responses are described as simultaneous pictorial and correct interpretations. From the total pictorial answers, the proportion of mixed answers was high: about the half of the verbalised pictorial/iconic responses were mixed. A detailed and carefully designed coding of the verbal interpretations revealed a complex intertwining of these students' conceptions. It was noted that some mixed cases showed a pictorial interpretation of a real or abstract object, or an abstract concept (e.g., a picture of time, a picture of noise) as well. In addition, it was found that the whole graph could be interpreted twice: once correctly and once pictorially; or part of the graph could be interpreted correctly and part of the graph could be interpreted pictorially. A statement of one of the participants exemplifies this idea; p35 interpreted Figure 5.11 as (the underlined text describes a gesture):

> *temperature... starts up quite high... and goes a bit down... it has been for that long* {<u>*points along the second part of the plotted line*</u>}

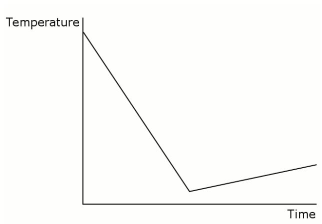


Figure 5.11: Line graph used in the Views battery.

There are two important consequences of these kind of pictorial answers. On the one hand, the "mixed" and the other pictorial interpretations could be characterised as "graph-as-a-picture misconceptions" under traditional classifications. However, this term seems inaccurate to describe the diverse states of the students' knowledge or to describe how close the student could be to achieving the expected conception. This was an important finding since it reinforces the idea that the students' pictorial interpretations are complex and the traditional graph-as-picture characterisation of their difficulties is too simple.

On the other hand, this diversity of interpretations could not have been observed by using only one of the batteries in this experiment. Each of the batteries (GCQ, GIQ, Views) seemed to capture a different snapshot of the students' knowledge. Thus, those tests had to be taken into account together in order to complement each other. In other words, different methods of assessment might be required to observe students' pieces of knowledge.

This result questions Berg and Smith's [11] result. They showed disparities on students' graphical abilities when using two different instruments, a drawing and a multiple-choice questionnaire. Particularly, when using the multiple-choice instrument, they found a higher proportion of the pictorial interpretations than when using a drawing test. Thus, they claimed a lack of validity when using multiple-choice instruments for looking at students' graphical abilities. However, their tests might have shown different students' interpretation knowledge, specifically because of the fact that both assessments were quite different (a construction task versus an interpretation task).

Overall, a rich set of tests might be suitable to better capture students' pieces of knowledge and differences in students' interpretations motivates us to revise the "graph-as-picture misconception" term.

Chapter 6

Discussion and conclusion

We pursued two objectives in this thesis. The first one was to investigate which medium for a learning environment is better to improve students' line graph interpretations. The second objective was to improve the way students' line graph interpretations are assessed. We discuss our results for each of those questions in the next two sections and we end this chapter with a summary of our general outcomes.

6.1 Testing two different mediums for improving line graph interpretation

The research question we investigated was:

> Will an active environment be more effective in improving students' interpretations of line graphs than a passive learning environment?

In order to investigate this question, we designed an active and a passive learning environment based around a Racing Car activity. The Racing Car activity was developed based on Janvier's paper-and-pencil tasks [38, 39] which, to some extent, have been influential in studies regarding line graph comprehension (e.g., [70], [45]). We improved Janvier's tasks based on (a) the literature about designing interactive learning environments to support students' understanding of abstractions (e.g., [70], [60]) and (b) our observations of students' behaviour (e.g., Chapter 3 and Chapter 4),

Both environments (active and passive) presented an activity where students could race a car (active mode) or watch how a car was raced (passive mode) along a track while its corresponding speed/distance line graph changed concurrently. Importantly, we used a yoked design: students assigned to the active mode were able to race the car at their will while pupils in the passive mode watched the replays of one of their peers who went through the active mode. Both external representations (ERs, track and graph) were dynamically linked, too (through *dynalinking*): the car's speed was reflected on a speedometer placed on top of the car, and the track and x-axis changed colour when the car was moved along the track, and the car's volume went up when the car's speed increased.

In experiment 1, the Racing Car activity and a Graph Comprehension Questionnaire were administered to students across the 3rd to 6th school years. There were some indicators of improved performance for the active group; however, it was possible that a ceiling effect could have been an alternative explanation of our test results. Thus, we decided to re-administer the activity to a new group of students. The intervention for experiment 2 consisted of the Racing Car activity and it was administered to pupils in year 5 only. In addition, we explored a new active mode, which allowed students to interact with the line graph instead of interacting with the car ("drawing the graph" condition). Although we could not confirm the results from experiment 1, we found out that our "drawing the graph" condition has potential to improve students' interpretations of line graphs (including those students who initially interpreted the graph pictorially).

6.1.1 Comparing this research with other works in the literature

Some research reports that after using active environments, students' interpretations of abstract representations improved. Nonetheless, some of those studies do not seem to compare students' experiences fairly between the active and passive modalities or constrain students to certain experiences (e.g., students have to follow a specific sequence of motions). For example, Anastopoulou, Sharples and Baber [3] concluded that the physical manipulation of graphs improved students' interpretations. However, their experiment compared two groups which had different experiences: their active group could manipulate the environment and benefit from (un)successful attempts during the task, but all the participants in their passive group observed a sequence of successful interactions performed by an experimenter only. Beichner's [9] groups differed on the tasks they were asked to do: the group under the "traditional instruction" performed a graph construction task, whereas the MBL group did not. In the case of Avons, Beveridge, Hickman and Hitch [5], students in the active mode had to act upon instructions from the experimenter instead of allowing pupils to explore the environment by themselves. In our study, however, the active modes were yoked with the passive modes; thus, we reduced the possibility of presenting our groups of students with very different experiences (in terms of the information displayed on the screen).

Note that most of the studies that focus on the benefits of MBL aim to improve pupils' knowledge on *curriculum subjects* such as Physics, Maths, Chemistry, etc. Nevertheless, in this thesis, our attention is towards improving students' knowledge of one of the *abstract representations* that are usually used as a means to learning a specific topic: line graphs. Other research has investigated how to improve students' knowledge of diagrams through technology (e.g., [1], [19], [70], [60]), but they have not investigated the type of medium that could be more effective to use – we addressed this issue here, though. In other words, as stated in Chapter 2, recognition of the importance of graphicacy has increased the research regarding how we can engage young students in activities that improve their knowledge of abstract representations, and our work contributes to this area.

6.1.2 The active_change_graph condition

Our results pointed out that certain types of medium (or certain *directed* degrees of interactivity) could be beneficial for improving students' interpretations of line graphs. Students who used the active_change_graph condition ("drawing the graph" condition) improved their performance (thus, their pictorial conceptions decreased) and this was the case for the pictorial group, too. We also noted that these results could not have been due to a test re-test effect since there was no significant improvement for the students in the other three groups. There are different aspects that could explain this tasks' efficacy and we discuss some of these below.

Dynamic generation of hypotheses

It is suspected that students' hypotheses were organised differently during the "drawing the graph" (active_change_graph) condition compared to the "race the car" (active_race_car) condition. For the "drawing the graph" condition, setting the configuration of the graph at the start of the activity might have forced students to form their hypotheses at the *start* of the task and to confirm/refute/change them during the replay; whereas the active_race_car condition might have produced several on-the-go hypotheses *during* the task, those hypotheses might not have been systematically organised and/or might not have been deeply processed. For example, p03, who first changed the graph and then watched the race, was asked what he was trying to do. He answered: *"I was trying to make sure it [would?] go around the bends slower."* P04, who raced the car along a track with bends, was asked what he was trying to do, too. His answer shows how he changed his goals during the task:

> My plan was to go around all the way quite fast and then I just decided to change the pattern because I felt like it ... I felt like going around all the way quite fast, but then I've got stuck so then I just decided to make that part of my pattern and go quicker a bit slower then a bit fast...

Our explanation can be complemented with Kaput's view [40] on translations between notations. Consider notation A as "the track" and notation B as "the graph":

> The process of relating actions in one notation A with actions or consequences in another related one B often proves cognitively overwhelming. In particular, one must become engaged in three different activities: (i) actions in A which effect state-changes in A; (ii) actions in B which effect state-changes in B; and (iii) coordinations of objects, relations, and most importantly, state-changes between A and B. Furthermore, given the limits on human cognitive processing capacity, these three activities must be performed serially in some order – they cannot occur simultaneously... ([40], p. 541).

Students in the active_race_car condition had to consider two activities simultaneously: (1) actions in the track (i.e. moving the car), and (2) state-changes between the track and graph. Whereas students in the active_change_graph condition had to consider two activities serially: (1) actions in the graph (i.e. changing the plotted line), and (2) state-changes between the track and the graph. Thus, the process of relating actions and consequences in the active_race_car condition might be more cognitively overwhelming than the process of relating actions and consequences in the active_change_graph condition.

Access to the task's main target

The fact that the active_race_car condition did not have a significant improvement in experiment 2 might seem to contradict other research. For instance, our results do not match with observations made by Beichner [9]. Beichner [9] claimed that MBL could be good at reducing pictorial conceptions if the graph and the event are linked together, but this claim now seems too broad in the light of our results. One of the main differences between Beichner's and our research lies in that his tasks were designed to improve knowledge of physics' concepts. The purpose of our environment is not about learning physics' concepts (where *motion within the learning environment* might play an important role in what students are learning), but we focused on the students' learning of the external representation instead (which might not necessarily be related/linked to a motion concept).

That is, in Beichner's activities, students seemed to have benefitted from the kinaesthetic feedback whereas in our tasks, continuous feedback was not as important as (1) gaining control of the ER, (2) giving students the experience of working with the representation, (3) letting them know (implicitly) some of the constraints of the ER (e.g., they were not able to change the axes during a race) and (4) converting the activity into a construction task. All those points seem more relevant to our purposes than giving continuous feedback from the hand motion.

Mixed interpretations

Mainly, students who went through the "drawing the graph" activity seemed to reduce their mixed conceptions, still this task had little effect on students whose conceptions where "pure pictorial" or "other"[1]. Thus, the active_change_graph condition seemed more effective for supporting those students whose knowledge might be in transition, and whose conceptions are close/moving towards the correct one. This means that the interaction with the graph might produce changes to pupils' knowledge structures and, although it encourages the transition towards a correct interpretation, it does not seem to encourage the creation/connection of useful pieces of knowledge for those students who tend to interpret the graph as a pure picture of the situation. The next steps could be taken in different directions to enrich pupils' experiences, and some of them could include: changing the task's context, giving more control of the ER, or observing how students work in groups.

6.1.3 Pictorial group

The pictorial group seemed to benefit from the active_change_graph condition too, but this condition had a greater effect on those students who elicited mixed interpretations than on those students who interpreted the graph as a "pure picture" of the situation. How could we improve line graph comprehension for those students who have pure pictorial conceptions? Based on Clement's [16, 17] work and our analyses, we could note that (1) when different chunks of the representation have different meanings (mixed interpretations) our "drawing the graph" condition seems useful, and (2) "pure pictorial" conceptions could be characterised by "global" interpretations of the ER (where the *whole* plotted line is

[1]By "other" we mean conceptions that were not correct, or mixed, or pure pictorial.

interpreted pictorially)[2]. Given this, perhaps one step forward could be to support the transition that takes students from the "pure" pictorial interpretation (where the plotted line is interpreted as one chunk) to a "mixed" interpretation (where the plotted line seems to be interpreted in different chunks).

Considering our research findings, it seems that interaction could be effective when the task focuses on the ER. Thus, we could suggest that the next direction to take could involve the refinement/development of an interactive learning environment that focuses the students' interaction on the ER and which tasks involve interpretation of the plotted line chunks. For example, the activity could include features that "force" the pupil to draw the plotted line by sections. We could show a toolbox with different plotted line sections (each section represents a function), the pupil could drag-and-drop or stretch sections in order to "draw" the plotted line (see a mock-up of this in Figure 6.1).

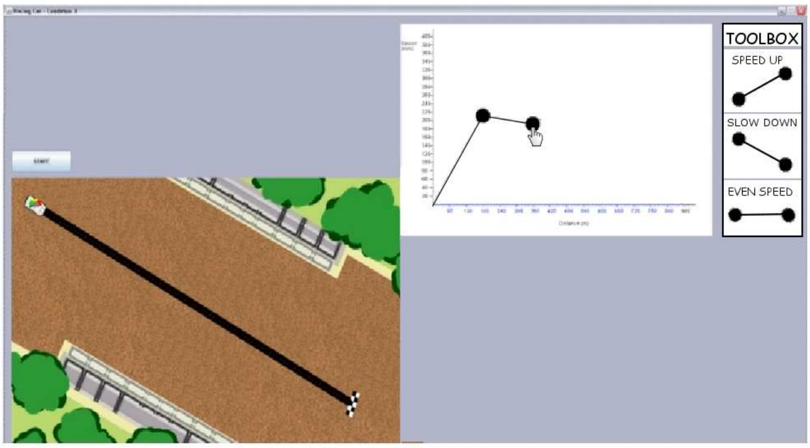


Figure 6.1: A toolbox with segments of the plotted line to drag-and-drop – the parts of the graph can be stretched.

6.1.4 Limitations and advantages of our approach

We are aware that some aspects of the task and method could have had an impact on our results. On the one hand, although we tried to pose neutral questions to students (e.g., *what were you trying to do? Which part were you looking at?*), we do not know if by asking them about their plans we unintentionally affected what students decided to do (this is similar to causing a self-explanation effect, e.g., [14]). On the other hand, we aimed to design a task that could fit within school activities, which means that group discussions between peers and/or with guidance of their teacher are more likely to take place; thus

[2] We cannot consider Clement's definition of a "local" pictorial interpretation as a "pure pictorial" conception since it is possible that other segments of the ER had been interpreted correctly by the student.

by questioning our participants we could have made our experiment more ecologically valid as well. Overall, a future step could be to investigate how to properly accommodate this task to school activities or how the student would behave if he/she was working by himself/herself.

Another important issue to consider is that our Racing Car tasks do not give the examiner/educator a clear cue/hint of the knowledge the student has just acquired. That is, the educator will need to rely on some kind of assessment in order to know how much progress the student has made (e.g., see Section 6.2).

This takes us to the next points: for the active modes, students had limited control of the ER. The benefit of this is that it allowed us to control certain aspects of the task (e.g., the part where the interaction happened), but we were not able to explore other pieces of students' knowledge related to other parts of the ER. For example, did pupils read the labels of the graph? Did they understand what the axes are for? What did they know about the scales of the graph?

We could also consider as an advantage of the active modes that the activities did not impose specific tasks (except for moving the car to the end of the track or modifying the graph). That is, those tasks allowed students to form their own plans/goals.

Overall, we could say that the active modes offered a "controlled" environment that allowed individual exploration, but at the same time they constrained what kinds of experiments were possible to perform. In spite of this, students seemed enthusiastic to perform these activities for both experiments.

During the passive modes, however, students were not able to control the animation and they had only one chance to replay it. The animations that were shown to students in the passive modalities might not have been the adequate medium to improve students' performance (e.g., [12]). However, in this experiment, it was necessary to avoid giving them extra control of the environment, otherwise comparison of the groups' outcomes would have been obscured by that extra variable.

6.1.5 Future work regarding the type of learning environment used for improving students' interpretations of line graphs

In this thesis we have mentioned some ideas that could be considered for further research. One line of work, however, seems to stand out for this author and it is about facilitating the interaction and construction of the graph – particularly, as discussed in Section 6.1.3, of the plotted line. That is, further research of the active_change_graph condition could be done to investigate more about how we can encourage the transition of students' knowledge, specifically for those students who have more difficulties to interpret the graph appropriately. For example, it would be interesting to find out whether by giving an explicit interpretation of parts of the plotted line (see Figure 6.1), students' graph comprehension improves and, if so, which strategies foment a better interpretation of the ER.

There are other aspects of the learning environment that could be tested (e.g., the level of "control" offered by the environment[3], using other/several background contexts). However, investigating which aspects of the active_change_graph condition improves young students' comprehension could help us to give advice on the construction of other diagrammatic interactive learning environments as well.

6.2 Assessing students' interpretations of line graphs

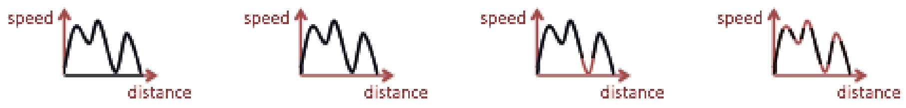
Our second research question was:

> How can we assess students' interpretations of line graphs?

We wanted to find an alternative method where we could observe various students' conceptions. Initially, we created a discrimination task with several items for the students to interpret, but pupils did not provide many verbal utterances that could help us to analyse their conceptions in depth. However, we noted that one of the tasks (the STQ in experiment 1) was very useful in spotting various types of interpretations and it was decided to include it (alongside other tests) as part of the assessment in experiment 2. Thus, in the second experiment we used several batteries as part of our assessment test, with this, we were able to (1) observe what kind of effect the learning environments was having on students' interpretations, and (2) ask several interpretation questions in different ways in order to observe and present more evidence of different types of (pictorial) conceptions.

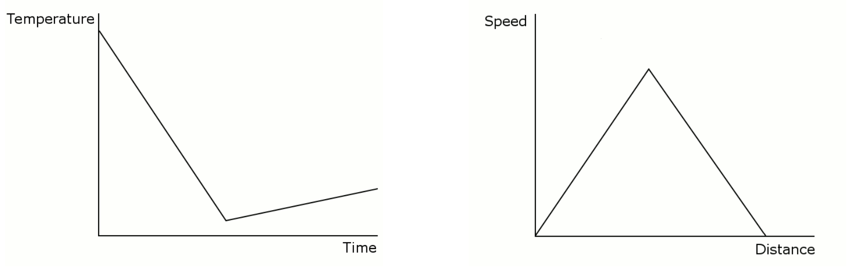
We have already used our rich set of tests to identify the "drawing the graph" condition as a promising learning environment to improve students' interpretations (see Section 5.5 and Section 6.1). In the next subsections we discuss a bit further the different types of pictorial conceptions we found, how can we explain them in terms of diSessa's framework

[3] We have used the term "controlled" in the sense that the interaction happens under specific conditions, defined by the types of tracks presented, the order of the tracks, the limited movements to change the graph or race the car, a forced replay, etc. Investigating which of those "controls" might or might not be necessary would be useful to improve the task.



(a) (b) (c) (d)

Figure 6.2: Examples of how different parts could have been attended in the plotted line of a line graph.



(a) (b)

Figure 6.3: Two line graphs.

(Section 6.2.1) and how those findings affect current (and previous) approaches of students' line graph assessment (Section 6.2.2).

6.2.1 Various forms of pictorial conceptions

In our empirical work we found evidence of various types of pictorial conceptions. This is a relevant finding since other research might have already suggested that some variations exist, but have not given further evidence of it (e.g., [53], [50]). In particular, we suggested that those variations could be distinguished by (1) how the student attends to different parts/segments/chunks of the graph and by (2) how each chunk is interpreted.

Let's consider Figure 6.2. Black coloured parts of the graphs represent different possible ways students could have attended to the graph. A student could have included the x-axis as part of the chunk used for interpretation (Figure 6.2a), whereas another student might have attended to the plotted line according to every change of its slope (Figure 6.2d). Our results suggest that not only "local" or "global" interpretations of the graph happen (as assumed by Clement [16]) but instead, a continuum exists between those extremes. That is, (a) the chunks students interpret could vary in size, (b) each chunk could be interpreted independently of the other chunks, and (c) each chunk could be interpreted several times; thus the way they are interpreted might not be only "global" or "local".

Furthermore, students' pictorial interpretations seemed more complex than was originally thought. We highlighted "mixed" types of interpretations, which are characterised by a partially correct interpretation and a pictorial interpretation of the graph (e.g., *"the* *person tried to go fast, but he was on a bumpy road"*, Figure 6.2c). We also noted that chunks could be interpreted as pictures of abstract concepts. For example, the line graph in Figure 6.3a could have been interpreted as[4]:

1. *"Time went down and up"*
   The graph is interpreted as a picture of an abstract concept: "time" (see Appendix G.2).

2. *"This is temperature {<u>he points along the first part of the plotted line</u>} and this is time {<u>he points along the second part of the plotted line</u>}"*
   Two abstract concepts are represented in different segments of the plotted line.

This result is relevant since it reveals the level of complexity of what has been previously described as a "picture of an object" or a "picture of the situation" only. In other words, the different pictorial conceptions do not match the traditional definition of the "graph-as-picture misconception". Nonetheless, these variations could be explained better using diSessa's framework [28].

Pictorial interpretations in terms of knowledge-in-pieces

We could consider Elby [29] and diSessa's [28] work to explain the variations of the pictorial interpretations that we found in our research. As we have mentioned in this work, those pictorial interpretations may be connected to the knowledge structure defined as WYSIWYG p-prim ([29], p. 488, see Section 2.4.2). Thus, we could suggest that (in terms of knowledge-in-pieces) when interpreting a representation:

a. Each individual could attend to certain chunks of the representation, which could be separated by compelling visual attributes (e.g., sharp changes of slopes).

b. Several knowledge structures would compete to interpret those chunks (and the WYSIWYG p-prim might be one of them).

c. According to the priority of each knowledge structure, some knowledge structures will "win" over others.

d. Chunks could be interpreted one time or more than one time.

For example, when interpreting the line graph in Figure 6.3b a person (Alex) could have (a) segmented the plotted line in two parts since the change of trend could have

[4]Descriptions of hand/finger gestures are underlined and within brackets.

worked as a compelling visual attribute for Alex. (b) The chunk / might have elicited a well-rehearsed knowledge structure that means “speeding up”. However, when the chunk \ is processed, the WYSIWYG intuitive knowledge structure could have had a higher (c) priority compared to other structures. Therefore, a pictorial interpretation could have been retrieved and the concept of “going down” could have been assigned to it.

6.2.2 Assessment

Noting those various forms of pictorial conceptions has important consequences on how we evaluate (and how other pieces of research that have evaluated) students’ line graph interpretations since this means that some pieces of research might not have accurately assessed students’ graphical knowledge before. An example of this is the research conducted by Berg and Smith [11]. They [11] observed disparities between multiple-choice and free-response instruments, but their results could have been affected by the different tasks they compared (graph interpretation vs. graph construction) and by how each instrument (administered to different students) triggered different aspects of pupils’ conceptions.

In this research, we noted that a richer set of tests would be more adequate for understanding students’ different pieces of knowledge since those might be elicited by different tasks. In addition, through a set of different tasks, evidence of other students’ conceptions could surface as well.

6.2.3 Limitations and advantages

Our outcomes point out that using a rich set of tests seems better to capture students’ pieces of knowledge than only one type of test. Nonetheless, this means that some of the batteries could be difficult to evaluate (such as the Views battery) than others (such as the MCQ). This could be a disadvantage for educators since some batteries might require more time to assess. A possible step forward is to investigate whether some methods are more effective than others in order to improve/facilitate evaluation tasks.

In terms of the application of the batteries in the school environment, some of them could be more challenging. Again, the Views battery asks the students to speak aloud instead of writing down their interpretations. Thus, it would be interesting to test if the Views battery could be replaced by a free-response test (although such a test might not be as rich – in terms of outputs – as the verbal/speak aloud option).

Another issue to consider is whether context will trigger certain knowledge structures, but not others. However, this is an issue that needs to be examined with a follow-up test.

6.2.4 Future work

As mentioned before, there are a few lines of investigation that could be pursued. For instance:

- We could investigate which combination of the instruments would provide the most effective approach? Do all contribute equally to see students' conceptions? Or are some of them more effective than the others?

- Additional questions we could explore are: What's the role of the context on the assessment batteries? Does it work the same across all contexts?

6.3 General outcomes

The goals of this research were to provide tools that could be used in school for observing students' graphical knowledge. Ideally, educators would be able to administer an assessment test in order to find out more about their students' graphical knowledge and to associate a curriculum activity to a learning environment that could help to improve students' understanding. Although there is still work to do in both areas, this work made contributions towards those goals.

Additionally, although this research was limited to improve students' interpretations of line graphs, our results could be extended to improve students' interpretations of other representational forms (e.g., Euler's circles, see Section 2.6.3). In particular, the design of the active_change_graph condition could be used to inform the design of other interactive learning environments that focus on students' learning of external representations.

Our research informs us that the traditional view of students' "graph-as-picture misconceptions" might not accurately reflect their students' knowledge. Instead, a rich set of tests might need to be administered in order to observe in more detail different students' pieces of knowledge. In other words, students' knowledge seems better evaluated when using a rich set of batteries and the instruments presented here could be used as the basis for this.

Additionally, active learning environments that permit exploration of the abstract representation may act as scaffolds for helping students to comprehend abstractions; thus, the "drawing the graph" condition has the potential to improve students' line graph interpretations. Borrowing the title of the article by Keehner, Hegarty, Cohen, Khooshabeh and Montello [41] we could say that for interactive learning environments designed for improving line graph interpretations: it does not matter *if* you interact, but with *what* you interact.

Bibliography

[1] J. Ainley. Building on children's intuitions about line graphs. In J.P. da Ponte and J.F. Matos, editors, Proceedings of the Eighteenth Annual Conference of the International Group for the Psychology of Mathematics Education, volume 2, pages 252–259, Lisbon, Portugal, 1994.

[2] F.K. Aldrich and L. Sheppard. Graphicacy: the fourth 'R'? Primary Science Review, pages 8–11, 2000.

[3] S. Anastopoulou, M. Sharples, and C. Baber. An evaluation of multimodal interactions with technology while learning science concept. British Journal of Educational Technology, pages 266–290, 2.

[4] R. Arnheim. Visual Thinking. Faber and Faber Limited, London, 1970.

[5] S.E. Avons, M.C. Beveridge, A.T. Hickman, and G.T. Hitch. Teaching journey graphs with microcomputer animation. Journal of Practical Research and Applications, 2(2):93–105, 1983.

[6] W.G.V. Balchin and A.M. Coleman. Graphicacy should be the fourth ace in the pack. The Times Educational Supplement, page 947, November 5 1965.

[7] W.L. Barclay. Graphing misconceptions and possible remedies using microcomputer-based labs. Technical report, Technical Education Research Center, Cambridge, Mass., November, 1985. ERIC Document Reproduction Service No. ED264129.

[8] S. Bateman, R. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In ACM Conference on Human Factors in Computing Systems (CHI 2010), pages 2573–2582, Atlanta, GA, USA, 2010. Best paper award.

[9] R.J. Beichner. The effect of simultaneous motion presentation and graph generation in a kinematics lab. Journal of Research in Science Teaching, 27(8):803–815, 1990.

[10] A. Bell, G. Brekke, and M. Swan. Diagnostic teaching: 4 graphical interpretation. Mathematics Teaching, 119:56–59, 1987.

[11] C.A. Berg and P. Smith. Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. Science Education, 78(6):527–554, 1994.

[12] M. Betrancourt. The animation and interactivity principles in multimedia learning. In R.E. Mayer, editor, The Cambridge handbook of multimedia learning, pages 287–296. Cambridge University Press, 2005.

[13] M.C. Beveridge, D.J. Dean, S.E. Avons, and G.J. Hitch. Teaching journey graphs to primary school children: A comparison of procedures. Educational Psychology, 3(2):149–159, 1983.

[14] M.T.H. Chi. Quantifying qualitative analyses of verbal data: A practical guide. The Journal of the Learning Sciences, 6(3):271–315, 1997.

[15] M.T.H. Chi. Active-constructive-interactive: A conceptual framework for differentiating learning activities. Topics in Cognitive Science, 1(1):73–105, 2009.

[16] J. Clement. Misconceptions in graphing. In Proceedings of the Ninth International Conference for the Psychology of Mathematics Education, pages 369–375, 1985.

[17] J. Clement, D.E. Brown, and A. Zietsman. Not all preconceptions are misconceptions: finding 'anchoring' conceptions for grounding instruction on students' intuitions. International Journal of Science Education, 11:554–565, 1989.

[18] J. Clement, J.R. Mokros, and K. Schultz. Adolescents' graphing skills: A descriptive analysis. Technical report, Technical Education Research Center, Cambridge, Mass., November, 1985. ERIC Document Reproduction Service No. ED264127.

[19] J. Corcoran. Review - data logging with 7- and 8-year-olds. Microscope, 32:45–49, 1991.

[20] R. Cox. Representation interpretation versus representation construction: A controlled study using switchERII. In B. du Boulay and R. Mizoguchi, editors, Artificial Intelligence in Education: Knowledge and media in learning systems. Proceedings of the 8th World Conference of the Artificial Intelligence in Education Society, pages 434–441, Amsterdam, 1997. IOS.

[21] R. Cox. Representation construction, externalised cognition and individual differences. Learning and Instruction, 9:343–363, 1999. Special issue on learning with interactive graphical systems.

[22] R. Cox. Technology-enhanced research: Educational ICT systems as research instruments. Technology, Pedagogy & Education, 16(3):337–356, 2007.

[23] R. Cox and P. Brna. Supporting the use of external representations in problem solving: The need for flexible learning environments. Journal of Artificial Intelligence in Education, 6(2):239–302, 1995.

[24] R. Cox and B. Grawemeyer. The mental organisation of external representations. In F. Schmalhofer, R. Young, and G. Katz, editors, European Cognitive Science Conference (EuroCogSci - joint Cognitive Science Society and German Cognitive Science Society conference). Proceedings of EuroCogSci03 (1st European Cognitive Science Conference), pages 91–96, Mahwah, NJ, 2003. Lawrence Erlbaum Associates.

[25] L. Cuban. Teachers and Machines: The Classroom Use of Technology Since 1920. Teachers College Press, New York, NY, 1986.

[26] F.R. Curcio and M.T. Smith-Burke. Processing information in graphical form. Technical report, American Educational Research Association, New York, NY, 1982. ERIC Document Reproduction Service No. ED215874.

[27] A.A. diSessa. Knowledge in pieces. In G. Forman and P. Pufall, editors, Constructivism in the computer age, pages 49–70. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1988.

[28] A.A. diSessa. Toward an epistemology of physics. Cognition and Instruction, 10(2-3):105–225, 1993.

[29] A. Elby. What students' learning of representations tells us about constructivism. The Journal of Mathematical Behavior, 19(4):481–502, 2000.

[30] K.A. Ericsson and H.A. Simon. Protocol analysis: Verbal reports as data (Revised edition). The MIT Press, Cambridge, MA, 1993.

[31] C. Fiolhais and J.A. Trindade. Use of computers in physics education. `http://nautilus.fis.uc.pt/personal/jtrindade/~jtrindade/pub/articles/art2/art2.html`, 2012. Retrieved on June 2012.

[32] S.N. Friel, F.R. Curcio, and G.W. Bright. Making sense of graphs: Critical factors influencing comprehension and instructional implications. Journal for Research in Mathematics Education, 32(2):124–158, 2001.

[33] G. Garcia Garcia and R. Cox. Diagrams in the UK national curriculum. In G. Stapleton, J. Howse, and J. Lee, editors, Diagrammatic Representation and Inference - Fifth International Conference on the Theory and Application of Diagrams 2008, LNCS 5223, pages 360–363, Herrsching, Germany, 2008. Springer.

[34] G. Garcia Garcia and R. Cox. An infra-red finger tracking system used in the assessment and remediation of "graph-as-picture" misconceptions. In R. Cox and J. San Diego, editors, Proceedings of the Workshop on Technology-Enhanced Diagrams Research, volume 887, pages 42–51, Canterbury, UK, 2012. CEUR Workshop Proceedings.

[35] D. Grayson and L.C. McDermott. Use of the computer for research on student thinking in physics. American Journal of Physics, 64(5):557–565, 1996.

[36] T.R. Greene. Children's understanding of class inclusion hierarchies: The relationship between external representation and task performance. Journal of Experimental Child Psychology, 48:62–89, 1989.

[37] G.J. Hitch, M.C. Beveridge, S.E. Avons, and A.T. Hickman. Effects of reference domain in children's comprehension of coordinate graphs. In D.R. Rogers and J.A. Sloboda, editors, The Acquisition of Symbolic Skills, pages 551–560, London & New York, 1983. Plenum Press.

[38] C. Janvier. The Interpretation of Complex Cartesian Graphs Representing Situations - Studies and Teaching Experiments. PhD thesis, University of Nottingham, 1978.

[39] C. Janvier. Use of situations in mathematics education. Educational Studies in Mathematics, 12(1):113–122, 1981.

[40] J. Kaput. Technology and mathematics education. In D. Grouws, editor, A handbook of research on mathematics teaching and learning, pages 515–556. Macmillan, 1992.

[41] M. Keehner, M. Hegarty, C. Cohen, P. Khooshabeh, and D.R. Montello. Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. Cognitive Science, 32(7):1099–1132, 2008.

[42] D. Kerslake. The understanding of graphs. Mathematics in School, 6(2):22–25, 1977.

[43] D. Kieras. A guide to GOMS model usability evaluation using NGOMSL, 1996. Retrieved September 2012 from http://www.cosc.brocku.ca/ bockusd/3p94/kieras96guide.pdf.

[44] M. Kozhevnikov, M.A. Motes, and M. Hegarty. Spatial visualization in physics problem solving. Cognitive Science, 31:549–579, 2007.

[45] G. Leinhardt, O. Zaslavsky, and M.K. Stein. Functions, graphs, and graphing: Tasks, learning, and teaching. Review of Educational Research, 60(1):1–64, 1990.

[46] P.D. Mautone and R.E. Mayer. Cognitive aids for guiding graph comprehension. Journal of Educational Psychology, 99:640–652, 2007.

[47] R.E. Mayer. Cognitive theory of multimedia learning. In R.E. Mayer, editor, The Cambridge handbook of multimedia learning, pages 31–48. Cambridge University Press, 2005.

[48] R.E. Mayer. Introduction to multimedia learning. In R.E. Mayer, editor, The Cambridge handbook of multimedia learning, pages 1–16. Cambridge University Press, 2005.

[49] L.C. McDermott. Research and computer-based instruction: Opportunity for interaction. American Journal of Physics, 58:452–462, 1990.

[50] L.C. McDermott, M.L. Rosenquist, and E.H. van Zee. Student difficulties in connecting graphs and physics: Examples from kinematics. American Journal of Physics, 55:503–513, 1987.

[51] D. McNeill. Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press, 1992.

[52] Z.R. Mevarech and B. Kramarsky. From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. Educational Studies in Mathematics, 32(3):229–263, 1997.

[53] J.R. Mokros and R.F. Tinker. The impact of microcomputer-based labs on children's ability to interpret graphs. Journal of Research in Science Teaching, 24(4):369–383, 1987.

[54] C. Morgan. Look at that big noise! Micromath, 8(2):10–13, 1992.

[55] C. Morgan. Where data handling meets information technology. can new resources enable children to interpret graphs more readily? a case study. Microscope, 38:19–21, 1993.

[56] T. Oppenheimer. The Flickering Mind: The False Promise of Technology in the Classroom and How Learning Can Be Saved. Random House, New York, NY, 2003.

[57] L. Pereira-Mendoza. Graphing in the primary school: Algorithm versus comprehension. Teaching Statistics, 17(1):2–6, 1995.

[58] R.J. Phillips. Can juniors read graphs? a review and analysis of some computer-based activities. Journal of Information Technology for Teacher Education, 6(1):49–58, 1997.

[59] F. Reif. Applying cognitive science to education: Thinking and learning in scientific and other complex domains. MIT Press, Cambridge, MA, 2008.

[60] Y. Rogers, M. Scaife, F. Aldrich, and S. Price. Improving children's understanding of formalisms through interacting with multimedia. Technical report, University of Sussex, 2003. CSRP Technical Report No. 559.

[61] J. Roschelle, N. Shechtman, D. Tatar, S. Hegedus, B. Hopkins, S. Empson, J. Knudsen, and L.P. Gallagher. Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. American Educational Research Journal, 47:833–878, 2010.

[62] J. Roschelle, D. Tatar, N. Shechtman, S. Hegedus, B. Hopkins, J. Knudsen, and A. Stroter. Scaling up SimCalc project: Can a technology enhanced curriculum improve student learning of important mathematics? Technical report, SRI International, 2007. Technical Report No. 01.

[63] M. Scaife and Y. Rogers. External cognition: how do graphical representations work? Int. J. Human-Computer Studies, 45(2):185–213, 1996.

[64] J.P. Smith, III, A.A. diSessa, and J. Roschelle. Misconceptions reconceived: A constructivist analysis of knowledge in transition. The Journal of the Learning Sciences, 3(2):115–163, 1994.

[65] J. Sweller. Implications of cognitive load theory for multimedia learning. In R.E. Mayer, editor, The Cambridge handbook of multimedia learning, pages 19–30. Cambridge University Press, 2005.

[66] K.L. Taylor and J-P. Dionne. Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. Journal of Educational Psychology, 92(3):413–425, 2000.

[67] E. Teubal, J. Dockrell, and L. Tolchinsky, editors. Notational Knowledge: Historical and Developmental Perspectives. Sense Publishers, Rotterdam, The Netherlands, 2007.

[68] F. Turner. Beginning teachers' use of representation. In British Society for Research into Learning Mathematics, volume 27, 2007.

[69] Y. Uesaka, E. Manalo, and S. Ichikawa. What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? Learning and Instruction, 17:322–335, 2007.

[70] M. Vogel, R. Girwidz, and J. Engel. Supplantation of mental operations on graphs. Computers & Education, 49(4):1287–1298, 2007.

[71] H. Wainer. A test of graphicacy in children. Applied Psychological Measurement, 4(3):331–340, 1980.

[72] H. Wainer. Understanding graphs and tables. Educational Researcher, 21(1):14–23, 1992.

[73] B.Y. White. ThinkerTools: Causal models, conceptual change, and science education. Cognition and Instruction, 10(1):1–100, 1993.

Appendix A

Pilot study: Card-sort items and multiple-choice questionnaire

A.1 Card-sort items

The items shown to the students for the pilot study (Chapter 3) are presented here. Although six representational types were used, we present the line graphs items only (see Table A.1). The cards were also classified in three parts (which reflect the order students are "expected" to sort them). More details about this task are in Section 3.1.2.

A.2 Multiple-choice questionnaire

The multiple-choice questionnaire used in the first pilot study is below. The correct answers to the questions is in bold and the graph used for all the questions is in Figure A.1.

1. What distance did Peter walk by 1 PM?

   1.5 MK

   0.5 KM

   1 KM

   2 KM

2. How long did Peter take to walk 3 km?

   1 hour

   2 hours

   1.5 hours

   3 hours

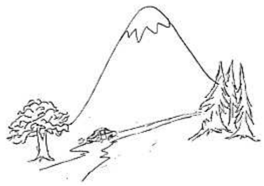
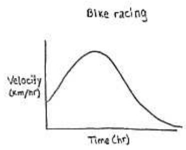
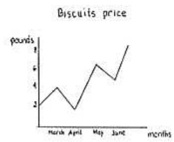
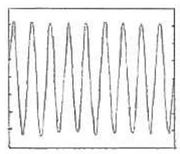
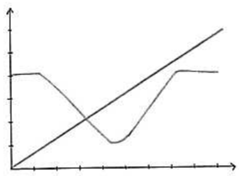
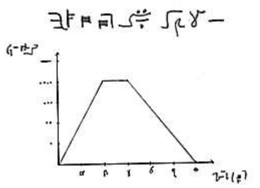
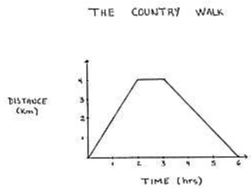
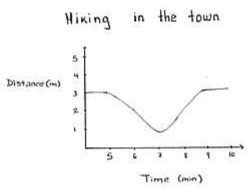
Order	Line Graph Items
Most	 
Little	   
None	  

Table A.1: Expected order for the cards for the cards showing line graphs.

3. How much did Peter walk between 3 PM and 5 PM?

   1 KM

   2 KM

   1.5 KM

   3 KM

4. Which of the following sentences describe what happened in Peter's trip?

   Peter walked more in the last 3 hours than in the first 3.

   Peter walked faster and then went down a hill.

   Peter walked up a hill and then went down the other side.

   Peter walked faster in the last 3 hours than in the first 3.

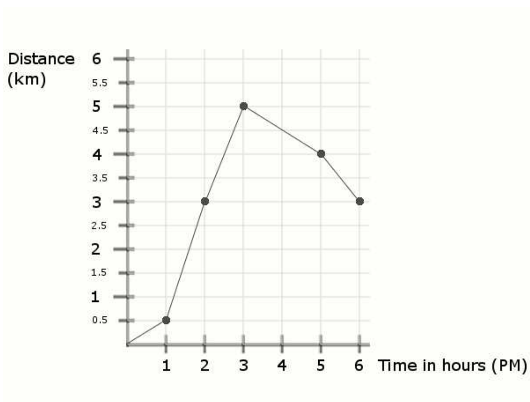
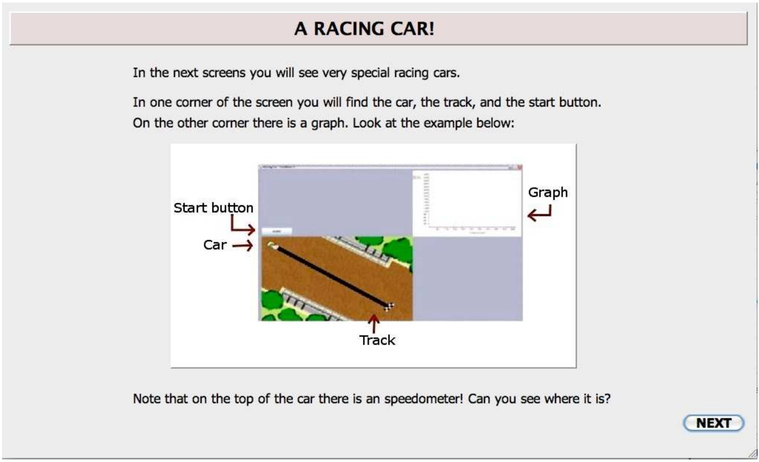


Figure A.1: Graph line used for the graph comprehension questionnaire.

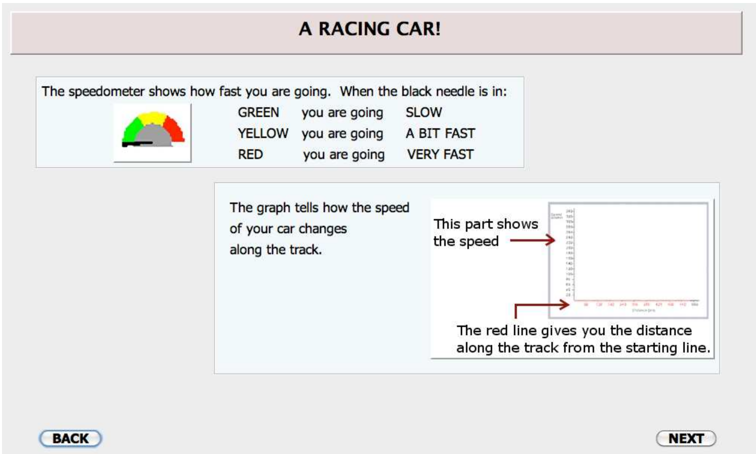
Appendix B

Experiment 1: Batteries

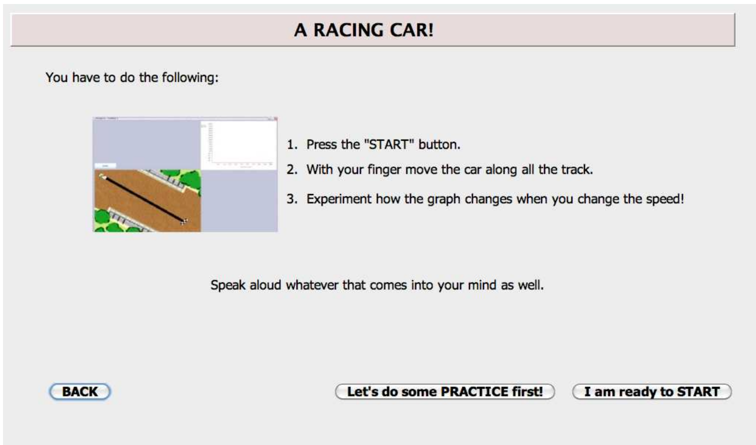
This part presents the batteries used in experiment one, Chapter 4. The Graph Comprehension questionnaire is in Table B.1; instructions for this questionnaire are on the top of the table. The items for the Discrimination Task are in Table B.2. Instructions for the Discrimination Task are in Section 4.4.2. The Racing Car activity instructions are presented in Figure B.1. Finally, the Select Track Questionnaire is in Table B.3; instructions for this questionnaire are on the top of that table.



(a) Instructions: Screen 1/3



(b) Instructions: Screen 2/3



(c) Instructions: Screen 3/3

Figure B.1: Screenshots of the Racing Car activity.

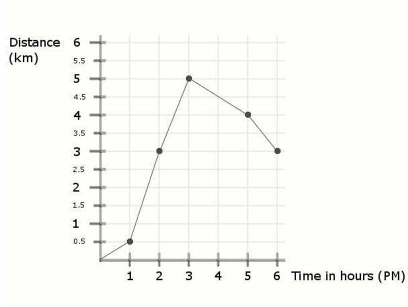
	Instructions. Peter is a kid that lives in a very very flat town. He likes to walk and he decided to go for a walk to town. The next screens show a graph about Peter's trip. Read the questions and answer them using the graph. You can take as much time as you need.
Line Graph	
Question 1: Reading x-axis	What distance did Peter walk by 1 PM? • 1.5 KM • 0.5 KM • 1 KM • 2 KM • Other answer
Question 2: Reading y-axis	For how long did Peter walk the first 3 km? • 1 hour • 3 hours • 5 hours • 2 hours • Other answer
Question 3: Operating two values	How much did Peter walk between 2 PM and 3 PM? • 1 KM • 2 KM • 1.5 KM • 3 KM • Other answer
Question 4: Interpreting	Which of the following sentences describe what happened in Peter's trip? • He walked slow in the first 3 hours and then he went fast. • He walked fast and then he went down a bridge. • He walks fast in the first 3 hours and then he walks slow. • He walks up a hill and then goes down the other side. • Other answer.

Table B.1: Graph Comprehension Questionnaire.

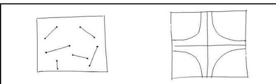
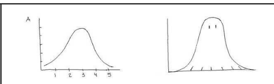
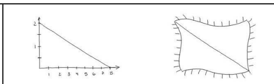
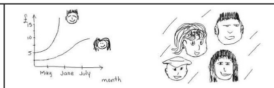
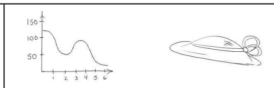
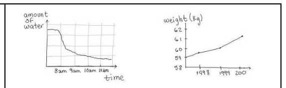
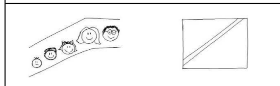
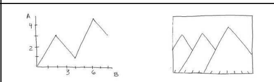
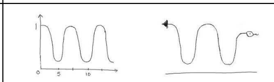
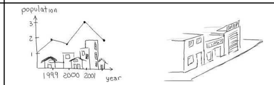
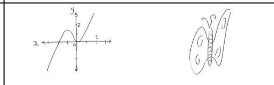
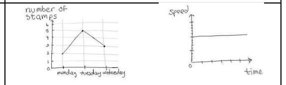
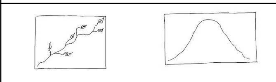
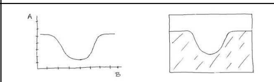
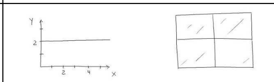
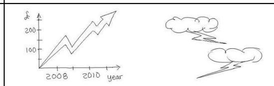
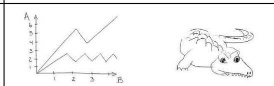
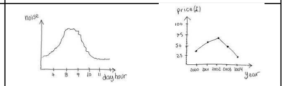
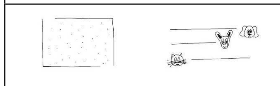
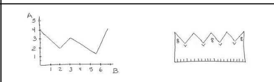
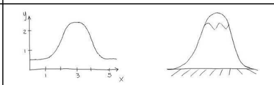
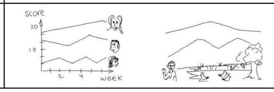
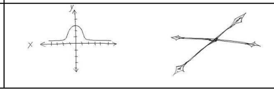
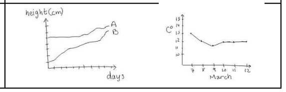
		Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6
		Picture, Picture	Diagram, Picture: small modification	Diagram, Picture: keep structure	Diagram with pictures, Picture	Diagram without pictures, Picture	Diagram, Diagram
Line graphs	Pre-test						
							
	Post-test						
							

Table B.2: Line graph items presented in the Discrimination Task.

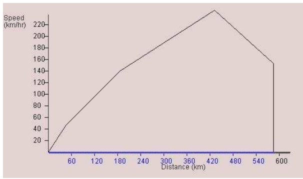
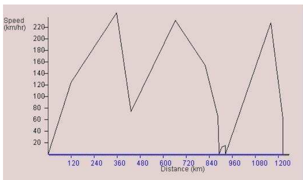
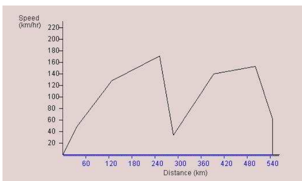
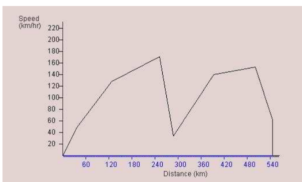
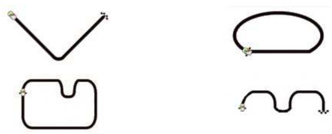
	Instructions. In the next screens you will see some graphs created by a smooth driver when doing the “Racing Car” activity. This time you will have to read the question and select the answer that you think is correct. You can take any time that you want. Ah! Remember that the tracks are scaled!!! That means that they are smaller on the screen than they are in reality!	
	Line Graph	Question and target items
Question 1:		Which track do you think corresponds to the graph?  other answer
Question 2:		Which track do you think corresponds to the graph?  other answer
Question 3:		Can you tell, from the graph, how many bends there are along the track on which the car was driven? • 1 • 2 • 3 • 5 • other answer
Question 4:		Which track do you think corresponds to the graph?  other answer

Table B.3: Select Track Questionnaire.

Appendix C

Experiment 1: Coding Scheme for the Select Track Questionnaire

This coding scheme is used on the utterances obtained for the Select Track Questionnaire from Experiment 1 (Chapter 4, Section 4.5.3). Its purpose is to identify pictorial utterances and describe students' graph interpretations.

C.1 General notes

Before giving a code to each of the utterances, we took care to identify each of the objects the student is referring to. That is, for each utterance we had to make sure that it was clear if the student was talking about or pointing to the speedometer, the graph, the track, etc. Part of this effort included using students' gestures (iconic/dietic). Thus, for instance, when the student pointed to an object then that gesture was considered as if the student had uttered/mentioned that specific object. Gestures were included in the transcript of the verbalisations and description of the gesture is in braces and underlined.

After several iterations for developing an appropriate coding scheme, 5 codes were used. The explanation of those codes is distributed along this appendix. The codes are: NO_UTTERANCE, AMBIGUOUS, PICTORIAL, NO_PICTORIAL and MIXED.

In additon, a list of key words was specified. Table C.1 shows some of the relevant concepts that the student might mention during the interpretation. This selection of words was used as tool to help us assigning the more adequate code to each of the students' verbalisations.

	Related to	Examples
Concepts:	Speed	fast, slow, medium speed, speed up, slow down, quick
	Distance	kilometers
Comparative concepts:	Speed	faster than, slower than
	Distance	longer than
Concepts as adjectives:	Speed	fast track, easy track, hard track
	Distance	long track
Color codes:	Speed	"green", "yellow" and "red" meaning slow, medium and fast speed respectively

Table C.1: Relevant concepts which could be mentioned during the graph interpretation.

C.2 Part 1: Identifying vague utterances

This part identifies vague utteranes. To proceed with the marking, look at the sections (below) that are indicated in the following "if-then" sentence:

```
If (``A'' does not apply) then
        if (``B'' does not apply) then
                CONTINUE MARKING: GO TO ``Part 2''.
        else
                MARK AS ``AMBIGUOUS''
        endif
else
        MARK AS ``NO_UTTERANCE''
endif
```

A. NO UTTERANCE The student does not give any utterance.

B. AMBIGUOUS If the utterance is similar to any of the cases below then it is marked as AMBIGUOUS. The vocabulary that can be used as a guide for marking the utterance is in bold.

- It is not possible to identify the referents or objects the student is talking about. Examples:

  "it went up and as it went down it slowed down"

  (The utterance might refer to the graph interpretation or to the car's behaviour on a track)

- ***"that one", "because", "do not know"***

  The student does not give any reason for his/her interpretation. That is, the student's response does not say why the specific answer was selected.

Examples:

"Probably that one"

"Why did you select that one? Because"

"I do not know"

- ***"feel"**, **"guess"***

  The student's reason does not give an insight into how the graph was interpreted. The utterance does not provide enough information about the interpretation.

  Examples:

  "Why did you select that one? I felt like"

  "it was a guess"

- ***"remember"**, **"did not do any"***

  The student does provide an utterance that it is not related to the interpretation of the "graph", "track" or "graph and track".

  Examples:

  "I remember I did that before"

  "We did not do any of them"

- *(agent) +* ***"go/goes/went/can go/has gone/get"*** *+ (description)*

  The student indicates that an agent (a person, a car, etc.) moves in certain way along the track (e.g., fast, quickly, straight, fastest, etc.) without mentioning or referring to the graph. The movement of the agent does not provide enough information about how the graph is interpreted.

  Examples:

  "it's gone quite fast around"

  "it goes {<u>hand moves in a "v" form: fast (from top to bottom), stop, fast(from bottom to top)</u>}"

  "it could go that fast"

  "he went a bit further"

  "it went straight"

  "it get the faster on that one"

C.3 Part 2: Description of the graph and track

This part identifies utterances that give descriptions of the track and/or the graph. To proceed with the marking, check the sections indicated in the following "if-then" sentence:

```
If (the student refers to the graph and does not refer to the track) then {
        If (the whole utterance is as described in ``C'' ) then
                MARK AS ``AMBIGUOUS''
        else
                CONTINUE MARKING: GO TO ``Part 3''.
        endif
} else {
        If (the student refers to the track and does not refer to the graph) then {
                SEE EXAMPLES OF UTTERANCES IN ``D''.
                MARK AS ``AMBIGUOUS''
        } else {
                If (the student refers to the track and the graph) then
                        CONTINUE MARKING: GO TO ``Part 3''
                else
                        MARK AS ``AMBIGUOUS''
                endif
        } endif
} endif
```

C. GRAPH DESCRIPTIONS These are graph descriptions which do not provide information about the student's interpretation. That is, the student might provide a description of the plotting without mentioning any relevant concept (examples of relevant concepts are in Table C.1). In that case, it is not possible to state if the pupil is able to understand the meaning of the abstract graphical representation. Therefore, the utterance should be marked as AMBIGUOUS.

- The student provides a description of the plotting without mentioning relevant concepts.

  Examples:

  "it has lot's of ups and downs"

  "It got a bit dodgy in the side"

- The student reads the axis (either x-axis or y-axis).

  Examples:

"it is 60 kilometers long"

"it went up to 200 kilometers per hour"

D. THE UTTERANCE ONLY REFERS TO THE TRACK Some examples of utterances that refer to the track are presented below. Note that the list of examples is not exhaustive.

- The student gives a description of the track.
  Examples:

  "it was a line"

  "it is a wobly track"

  "it's one bend"

  "it has lot's of bends"

  "it has lot's of ups and downs"

- The student characterizes the track by giving it an adjective.
  Examples:

  "fast track"

  "hard track"

- The student gives a description of the track and how the agent moves along the track (fast, slow, stop, etc.).
  Examples:

  "there's lots of bends, you can't go really that fast"

  "it was a straight line so it probably went a bit fast"

C.4 Part 3: Interpreting the Graph

This part reviews students' utterances that refer to the graph. To proceed with the marking, check the sections indicated in the following "if-then" sentence:

```
If if(part/all the description of the utterance is in ``E'')  then {
        if(part of the description is in ``F'') then
                MARK AS ``MIXED''
        else
```

```
                MARK AS ``PICTORIAL''
        endif
} else {
        if(part/all the description of the utterance is in ``F'') then
                MARK AS ``NO_PICTORIAL''
        else
                MARK AS ``AMBIGUOUS''
        endif
} endif
```

E. PICTORIAL INTERPRETATION When the whole/part of a graph is interpreted as an object (real world objects as: roads, mountains), that is, parts/all the graph sections are not given the corresponding abstract meaning, but they are assigned a literal/iconic correspondence to an object it is said a pictorial interpretation happens.

For this questionnaire, the graph is the abstract representation and the objects are the tracks presented to the students. Note that the student might suggest any other real world object. An example of a pictorial interpretation is when parts of the graph-plotting are identified/mapped/assigned as parts of the track, then the plotting becomes the track, the plotting is the track.

For the next student's interpretation, if there is a way to change "it", "that", etc., to the referent (graph/object) and the resulting sentence(s) indicates that the reason to select the particular track is because the graph has the same/very similar visual features than the object selected then the utterance should be marked as "pictorial". Examples of the vocabulary to look at are next.

- *"X **goes like** Y"*

  Where X is the graph and Y is the object or Y is the graph and X is the object.

  Examples:

  > *"I think it might be that 'cause that sort of goes [like] the track a bit"*
  >
  > *"it goes up and down {<u>hand imitates the track</u>} just like that one"*

- *" X is **the same as** Y"*

  Where X is the graph and Y is the object or Y is the graph and X is the object.

  Examples:

  > *"that one {<u>points to the track</u>} because, like it's the same as if <UM> when it's higher up {<u>moves finger along the plotting line</u>}..."*

"because it's exactly the same"

- *"X **looks like** Y"*

  Where X is the graph and Y is the object or Y is the graph and X is the object.

  Examples:

  "that graph sort of looks a bit like the track"

  "it kind of looks like it, because it looks like that {points to the track} compared to these {points to other tracks}"

- *"**t**hat is that one"* Where "that" is the graph and "one" is the object or "one" is the graph and "that" is the object.

  Examples:

  "...that's got bumps in it {points to the track} so it's that one {points to the graph}..."

F. NON PICTORIAL INTERPRETATION If the student assigns one or more of the relevant concepts (see Table C.1) to the whole or parts of the plotted line, then the student is not making a pictorial/iconic interpretation of the graph.

In addtion, the student might identify parts of the graph, assign one or more relevant concepts to them and match those parts of the graph and concepts to specific sections of the track (bends, straight segments). Conversely, the student might match track-sections to specific concepts and to specific parts of the graph. For example, the pupil might have implied that in straight segments of the track the car goes fast and in bends the car slows down and (as consequence) the plotting will be "up" and "down", meaning, "fast" and "slow".

Although the best scenario is when the student assigns the concepts to the appropriate graph features, it could be the case that the student assigns one or more concepts to parts of the graph/plotted line that are innapropriate (e.g. interpreting a "stop" as "going slow"). However, this coding is focused on identifying pictorial interepretations rather than accurate answers. Therefore, an answer that correctly/incorrectly assigns one or more relevant concepts to the graph is not considered a pictorial interpretation.

Vocabulary to look at when only the graph is interpreted:

- *"X there"*

  Where X is one of the relevant concepts and "there" refers to the graph. The student might point to the graph instead of mentioning "there".

Examples:

> *"It went fast there..."*
>
> *"...maybe it stop there..."*
>
> *"if I just speed on it {points to the plotted line}"*

- *"X, Y, Z,... there"*

  Where X, Y, Z are relevant concepts refering to the plotted line. That is, the student gives a conceptual description of the plotted line. The student might point to the graph instead of mentioning "there".

  Example:

  > *"slower and then it's fast and then it's really fast"*

- *"(description of the plotted line) + (concept)"*

  That is, one of the relevant concepts is part of a description of the plotted line showing that part of the graph is being interpreted appropriately.

  Examples:

  > *"up, down, up, down, stop..."*
  >
  > *"up, down, slow down a bit..."*

- *"(description of the plotted line) + (agent) +* ***goes*** *+ (concept)"*

  The student assigns a concept to an specific feature of the plotted line.

  Examples:

  > *"it goes down, it has to be slower"*
  >
  > *"...m-shape, so it goes fast..."*
  >
  > *"it only had one down bit and that it's when it goes slow"*

Vocabulary to look at when the graph and track are interpreted together.

- ***"there is*** *X" + (parts-in-the-track) + "and there is X" + (parts-in-the-graph)*

  Where X is a number and "parts-in-track" refers to either the sides or the bends of the track. The order of the sentence is not fixed and it might start with "parts-in-graph" first.

  Examples:

"there's three bumps... maybe... four {points to the track}, because there's fourth bits {points to the graph}"

"... there's three and there's three sides there."

- *[(track description) + (agent) + "can/cannot/goes" + (concept)] + (assertion using concept obtained from the graph)*

  The statement has three parts: (1) a description of the track; (2) how the agent moves along the described part; and (3) a sentence of how the agent should move (this concept must have been obtained from the graph). Note that the order of the sentences is not important as soon as two conditions are met: (a) There is a clear consequence between the track description and the movement of the agent; and (b) the asertion of the concept is obtained from the graph.

  Example 1:

  $$\text{“}\underbrace{\textit{there's lots of bends,}}_{\textit{track description}}\ \underbrace{\textit{you can't go really that fast}}_{\textit{agent + cannot go + concept}}\ \textit{and}\ \underbrace{\textit{this is going really fast.”}}_{\textit{asertion using concept}}$$

  Example 2:

  $$\text{“}\underbrace{\textit{it doesn't have round corners}}_{\textit{track description}}\textit{...}\ \underbrace{\textit{it [would] slow around the corners,}}_{\textit{agent + cannot go + concept}}$$

  $$\textit{but}\ \underbrace{\textit{that one was really fast}}_{\textit{asertion using concept}}\text{”}$$

Appendix D

Experiment 1: Task Analysis

The following GOMS-like formalism provides an idealised model of the process for solving each question presented in the Select Track Questionnaire. See Section 4.5.5.

It is worth to recall that the model (1) represents only a fraction of the many ways to answer the problem and (2) it might be unrealistic (e.g., it imposes an ordered sequence of steps). For instance, not everybody will first identify *all* the track sections and then assign a concept to each section of the track (lines 34 and 37). However, the aim of this model is to better understand how pictorial interpretations could occur.

In spite of the drawbacks of the formalism, the model is very helpful to identify, communicate and refer to student's knowledge and their interpretations.

A version of the GOMS model [43] was used as a base for the task analysis. The GOMS model helps to represent the procedures to solve the task. In particular, the operators that are used to specify the "knowledge of the situation" are *identify* and *assign*. The *identify* operator specifies how an external representation is chunked into units. Whereas the *assign* operator relates each of the chunked units of the external representation to a conceptual value. For instance, the different parts of a graph are identified in lines 137 to 151, but the concept assigned to them is expresed in lines 113 to 118.

A description of the notation used is explained below. Text in $<$ and $>$ is mandatory; a plus sign (+) means "at least one of" (e.g., `STATEMENT+` means "at least one STATEMENT"); "::=" means "is defined as"; and "|" means "or".

```
STATEMENT ::= GOAL | METHOD | SELECT | OPERATOR
OPERATOR  ::= GOTO | KEEP-RECORD-OPERATOR | CONDITION | FOR-OPERATOR | IF-OPERATOR
            | ASSIGN | IDENTIFY

GOAL:<GOAL'S NAME>  (REQUIRES requirement+)
```

```
        STATEMENT+   |   ENDGOAL.
ENDGOAL

METHOD: <METHOD'S NAME>  (REQUIRES requirement+)  (ASSUMPTION assumption+)
        STATEMENT+   |   ENDMETHOD.
ENDMETHOD

SELECT:
        METHOD+
ENDSELECT

ASSIGN <DESCRIPTION>
        IF-OPERATOR+
ENDASSIGN

IDENTIFY <DESCRIPTION> <REQUIRES requirement>
        IF-OPERATOR+   |   ENDIDENTIFY.
ENDIDENTIFY

GOTO:  GOAL   |   IDENTIFY   |   ASSIGN

KEEP-RECORD-OPERATOR ::= [RETAIN list-of-things-to-remember]

CONDITION    ::=  <description of the objects to retrieve from memory>

FOR-OPERATOR ::= For each [CONDITION]
                        STATEMENTS+
                ENDFOR

IF-OPERATOR ::= If [CONDITION] Then
                        STATEMENTS+
                else
                        STATEMENTS+
                ENDIF

IF-OPERATOR ::= If [CONDITION] Then
                        STATEMENTS+
                ENDIF

assumption  ::=  <description of how the objects are retrieved>

requirement ::=  CONDITION
```

Note that ENDMETHOD or ENDGOAL are used to quit the method or the goal, respectively. In addition, the operators ASSIGN and IDENTIFY are in blue. Some steps have a double red star (**). The stars are used to indicate the statements where pictorial interpretations might occurr. A double plus sign (++) indicates possible places where the picture could be interpreted as a graph. The task starts with the goal Select_Track_That_Corresponds_To_The_Graph (lines 1 to 5).

```
1 GOAL: Select_Track_That_Corresponds_To_The_Graph
2         SELECT: METHOD: INTERPRET_GRAPH_THEN_OBSERVE_TRACKS
3                 METHOD: OBSERVE_TRACKS_THEN_INTERPRET_GRAPH
4         ENDSELECT
5 ENDGOAL

6 GOAL: Interpret_Information_From_Graph
7         Recognize graph as a graph **
8         GOAL: Identify_Plotting_From_Graph
9                 GOTO: IDENTIFY graph sections **
10                 For each graph section 'x':
11                         Select GRAPH-SECTION 'x'
12                         GOTO: ASSIGN GRAPH-CONCEPT to GRAPH-SECTION('x') **
13                         If GRAPH-CONCEPT is 'PLOTTING' Then
14                                 [RETAIN 'x' is 'PLOTTING']
15                                 ENDGOAL.
16                         else
17                                 Continue with the next GRAPH-SECTION
18                         ENDIF
19                 ENDFOR
20         ENDGOAL
21         GOAL: Retrieve_Semantic_Information_For_Plotting_Sections
22                 Look at the plotting
23                 GOTO: IDENTIFY plotting sections **
24                 For each section 'x' of the plotting:
25                         Select PLOTTING-SECTION 'x'
26                         GOTO: ASSIGN MOVEMENT-CONCEPT to PLOTTING-SECTION('x') **
27                 ENDFOR
28                 [RETAIN OF THE "PLOTTING-MOVEMENT-CONCEPTS"]
29         ENDGOAL
30 ENDGOAL
```

```
GOAL: Obtain_A_Sequence_Of_The_Agent's_Movements_For_One_Track
        Recognize the option as a track **
        Recognize car starting point ++
        GOTO: IDENTIFY track sections **
        For each section 'y' of the track:
                Select TRACK-SECTION 'y'
                GOTO: ASSIGN MOVEMENT-CONCEPT to TRACK-SECTION('y') **
        ENDFOR
       [RETAIN OF THE "TRACK-MOVEMENT-CONCEPTS"]
ENDGOAL

METHOD: INTERPRET_GRAPH_THEN_OBSERVE_TRACKS
        GOTO: GOAL: Interpret_Information_From_Graph
        [RETAIN OF "PLOTTING-MOVEMENT-CONCEPTS"]
        For each option 'z' of the possible answers:
                Select option 'z' **
                GOTO: GOAL: Obtain_A_Sequence_Of_The_Agent's_Movements_For_One_Track('z')
                [RETAIN OF "TRACK-MOVEMENT-CONCEPTS"]
                SELECT: METHOD: PLOTTING-TRACK-MATCHING
                        METHOD: TRACK-PLOTTING-MATCHING
                ENDSELECT
                [RETAIN OF "GOAL-MATCH-IS-SATISFIED" OR "GOAL-MATCH-IS-NOT-SATISFIED"]
                If "GOAL-MATCH-IS-SATISFIED" Then
                        [KEEP-RECOD OPTION 'z' IS THE ANSWER]
                        ENDMETHOD.
                Else
                        Continue with the next possible answer.
                ENDIF
        ENDFOR
        [RETAIN "OTHER-ANSWER"]
ENDMETHOD

METHOD: OBSERVE_TRACKS_THEN_INTERPRET_GRAPH
        For each option 'z' of the possible answers:
                Select option 'z' **
                GOTO: GOAL: Obtain_A_Sequence_Of_The_Agent's_Movements_For_One_Track
                [RETAIN OF "TRACK-MOVEMENT-CONCEPTS"]
                If there-is-not "PLOTTING-MOVEMENT-CONCEPTS" Then
                        GOTO: GOAL: Interpret_Information_From_Graph
                        [RETAIN OF "PLOTTING-MOVEMENT-CONCEPTS"]
                ENDIF
                SELECT: METHOD: PLOTTING-TRACK-MATCHING
```

```
                        METHOD: TRACK-PLOTTING-MATCHING
                ENDSELECT
                [RETAIN OF "GOAL-MATCH-IS-SATISFIED" OR "GOAL-MATCH-IS-NOT-SATISFIED"]
                If "GOAL-MATCH-IS-SATISFIED" Then
                        [KEEP-RECOD OPTION 'z' IS THE ANSWER]
                        ENDMETHOD.
                Else
                        Continue with the next possible answer.
                ENDIF
        ENDFOR
        [RETAIN "OTHER-ANSWER"]
ENDMETHOD

METHOD: PLOTTING-TRACK-MATCHING        REQUIRES: TRACK-MOVEMENT-CONCEPTS
                                                 PLOTTING-MOVEMENT-CONCEPTS
                                                 ASSUMPTION: THE LISTS ABOVE ARE ORDERED

        For each PLOTTING-MOVEMENT-CONCEPT 'pmc' with index 'i'
                Obtain the TRACK-MOVEMENT-CONCEPT 'tmc' with index 'i'
                If 'pmc' = 'tmc' Then
                        Check PLOTTING-MOVEMENT-CONCEPT with index 'i+1'
                Else
                        [RETAIN "GOAL-MATCH-IS-NOT-SATISFIED"]
                        FINISH METHOD.
                ENDIF
        ENDFOR
        [RETAIN "GOAL-MATCH-IS-SATISFIED"]
ENDMETHOD

METHOD: TRACK-PLOTTING-MATCHING        REQUIRES: TRACK-MOVEMENT-CONCEPTS
                                                 PLOTTING-MOVEMENT-CONCEPTS
                                                 ASSUMPTION: THE LISTS ABOVE ARE ORDERED

        For each TRACK-MOVEMENT-CONCEPT 'tmc' with index 'i'
                Obtain the PLOTTING-MOVEMENT-CONCEPT with index 'i': 'pmc'
                If 'tmc' = 'pmc' Then
                        Check TRACK-MOVEMENT-CONCEPT with index 'i+1'
                Else
                        [RETAIN "GOAL-MATCH-IS-NOT-SATISFIED"]
                        FINISH METHOD.
                ENDIF
        ENDFOR
```

```
        [RETAIN "GOAL-MATCH-IS-SATISFIED"]
ENDMETHOD

ASSIGN GRAPH-CONCEPT to GRAPH-SECTION('x')
        If 'x' =  Then [RETAIN 'X-AXIS']
        If 'x' =  Then [RETAIN 'Y-AXIS']
        If 'x' = STRING* Then [RETAIN 'LABELS']
        If 'x' =  Then [RETAIN 'PLOTTING'] **
ENDASSIGN

ASSIGN MOVEMENT-CONCEPT to PLOTTING-SECTION('x')
        If 'x' =  Then [RETAIN 'SPEED-UP'] **
        If 'x' =  Then [RETAIN 'SLOW-DOWN'] **
        If 'x' =  Then [RETAIN 'CONSTANT']
        If 'x' =  Then [RETAIN 'SLOW-DOWN AND SPEED-UP']
        If 'x' =  Then [RETAIN 'SPEED-UP AND SLOW-DOWN']
        If 'x' =  AND  Then [RETAIN 'STOP']
        If 'x' =  AND  Then [RETAIN 'MAXIMUM-SPEED']
ENDASSIGN

ASSIGN MOVEMENT-CONCEPT to TRACK-SECTION('y')
        If 'y' =  Then [RETAIN 'FAST']
        If 'y' =  Then [RETAIN 'FAST'] **
        If 'y' =  Then [RETAIN 'FAST'] **
        If 'y' =  Then [RETAIN 'SLOW']
        If 'y' =  Then [RETAIN 'SLOW']
        If 'y' =  Then [RETAIN 'SLOW']
        If 'y' = Then [RETAIN 'SLOW']
ENDASSIGN

IDENTIFY graph-sections        REQUIRES: GRAPH-SEGMENT('s')

        If 's' = STRING* Then [RETAIN 's'] ENDIDENTIFY. ENDIF
        If 's' =  Then [RETAIN 's'] ENDIDENTIFY. ENDIF
        If 's' =  Then [RETAIN 's'] ENDIDENTIFY. ENDIF
        If 's' =  Then [RETAIN 's'] ENDIDENTIFY. ENDIF
        If 's' =  Then
                GOTO: IDENTIFY graph-sections for section
                GOTO: IDENTIFY graph-sections for section
        ENDIF
        If 's' =  Then
                GOTO: IDENTIFY graph-sections for section
```

```
                    GOTO: IDENTIFY graph-sections for section
        ENDIF
ENDIDENTIFY

IDENTIFY plotting sections          REQUIRES: PLOTTING-SEGMENT(‘s’)

        If ‘s’ has zero corners Then
                If ‘s’ =   Then [RETAIN ‘s’] ENDIDENTIFY. ENDIF
                If ‘s’ =   Then [RETAIN ‘s’] ENDIDENTIFY. ENDIF
                If ‘s’ =   Then [RETAIN ‘s’] ENDIDENTIFY. ENDIF
        Else
                If ‘s’ is one corner Then
                        If ‘s’ =   Then [RETAIN ‘s’] ENDIDENTIFY. ENDIF
                        If ‘s’ =   Then [RETAIN ‘s’] ENDIDENTIFY. ENDIF
                Else
                        Select first corner ‘c1’
                        GOTO: IDENTIFY plotting sections for ‘c1’
                        GOTO: IDENTIFY plotting sections for section previous to ‘c1’
                        GOTO: IDENTIFY plotting sections for section after ‘c1’
                ENDIF
        ENDIF
ENDIDENTIFY

IDENTIFY track sections          REQUIRES: TRACK-SCECTION(‘t’)

        If ‘t’ has zero coners Then
                If ‘t’ =   Then [RETAIN ‘t’] ENDIDENTIFY. ENDIF
                If ‘t’ =   Then [RETAIN ‘t’] ENDIDENTIFY. ENDIF
                If ‘t’ =   Then [RETAIN ‘t’] ENDIDENTIFY. ENDIF
        Else
                If ‘s’ is one corner Then
                        If ‘s’ =   Then [RETAIN ‘t’] ENDIDENTIFY. ENDIF
                        If ‘s’ =   Then [RETAIN ‘t’] ENDIDENTIFY. ENDIF
                        If ‘s’ =   Then [RETAIN ‘t’] ENDIDENTIFY. ENDIF
                        If ‘s’ =   Then [RETAIN ‘t’] ENDIDENTIFY. ENDIF
                Else
                        Select first corner ‘c1’
                        GOTO: IDENTIFY track sections for ‘c1’
                        GOTO: IDENTIFY track sections for section previous to ‘c1’
                        GOTO: IDENTIFY track sections for section after ‘c1’
                ENDIF
        ENDIF
```

```
ENDIDENTIFY
```

Appendix E

Experiment 2 - Batteries and Instructions

The batteries presented next are the pre-test and post-test of experiment 2. Details of the design of these batteries can be found in Section 5.3.

Table E.1 shows the Graph Comprehension Questionnaire (GCQ). The table contains the batteries used in the pre-test (Battery A) and post-test (Battery B). The instructions presented to participants are on the top of the table. Note that each screen presented the graph on the left hand side and its corresponding question and distracters on the right hand side (Figure E.1, top-left).

Table E.2 presents the Graph Interpretation Questionnaire (GIQ, top-right, Figure E.1). The pre-test and post-test are presented on the left and right side of the table. Note that, unlike the GCQ, each question shows different graphs. The sequence of items presented to the participants is from top to bottom.

The Graph and Tracks Questionnaire (GTQ) is in Table E.3. Instructions are on the top of the table and each question contains different graphs and tracks. The sequence presented to the participant can be read from top to bottom. See Figure E.1, bottom-left.

The last battery is shown in Table E.4. The student was required to give an interpretation for each of the items presented. The prompting question for each item was *"what can you tell me about this?"*. See Figure E.1, bottom-right.

Instructions for the active_race_car and active_change_graph conditions of the Racing Car activity are presented in Figure E.2 and Figure E.3, respectively.

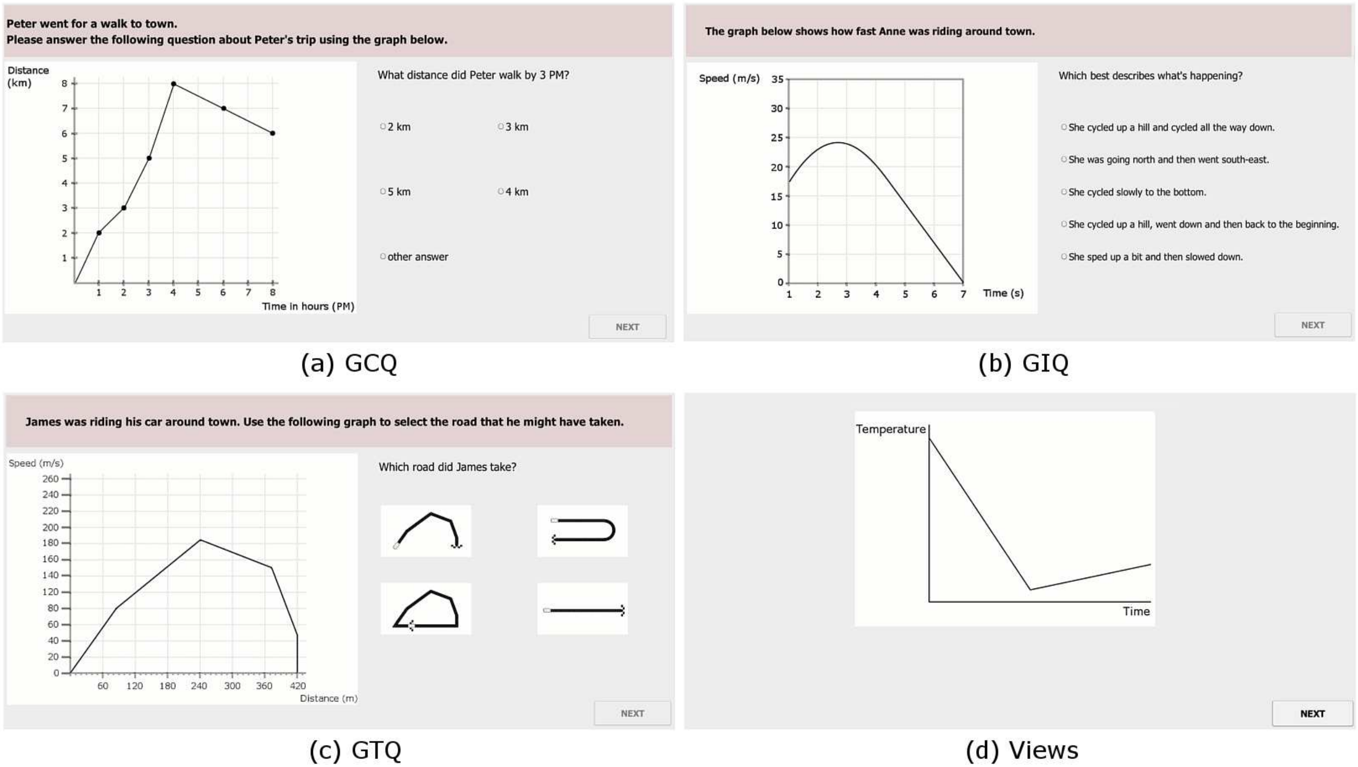
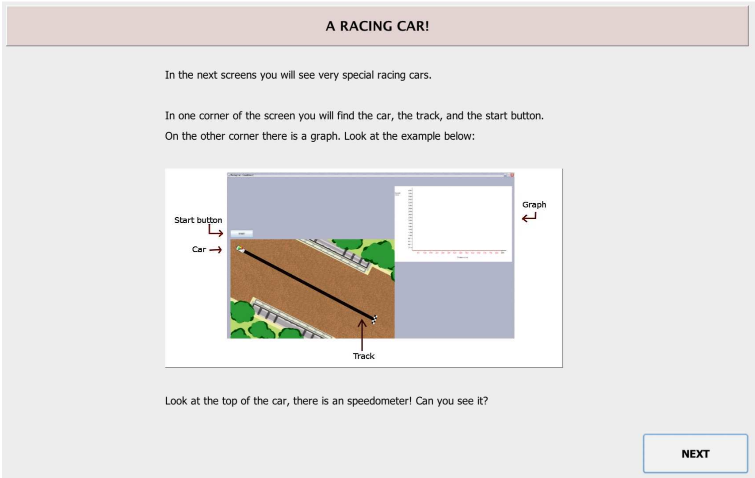
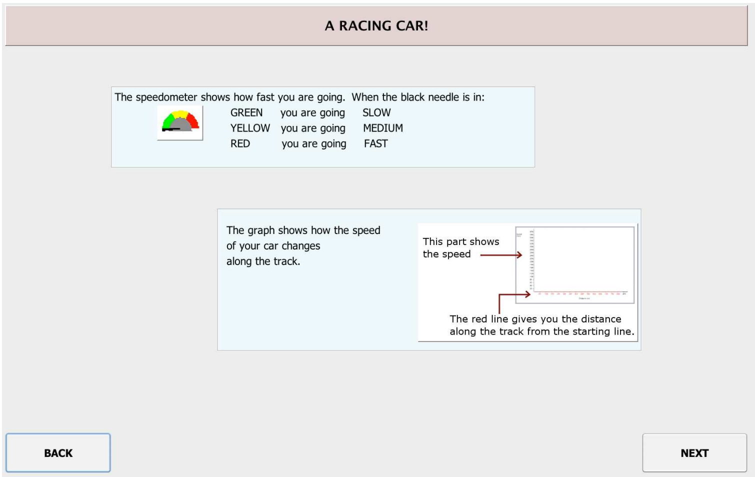


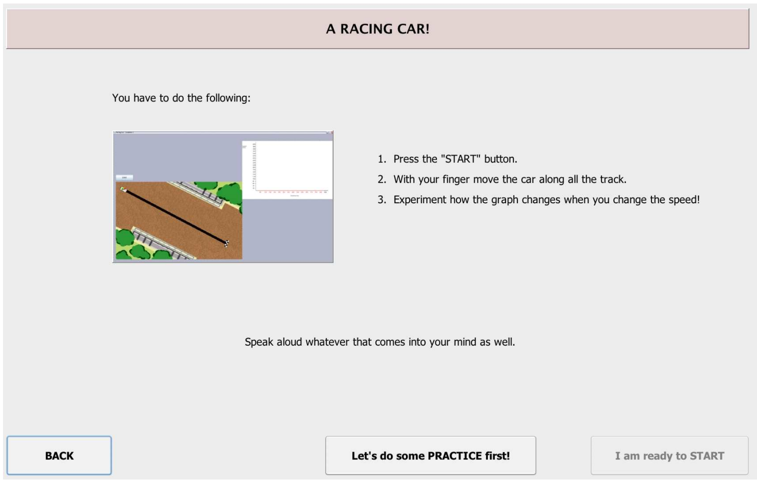
Figure E.1: Screenshots of the batteries for pre/post-test of experiment 2.



(a) Instructions: Screen 1/3

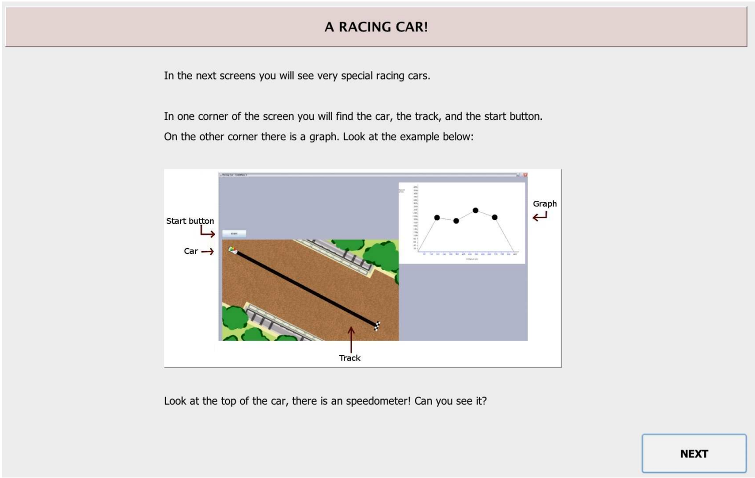


(b) Instructions: Screen 2/3

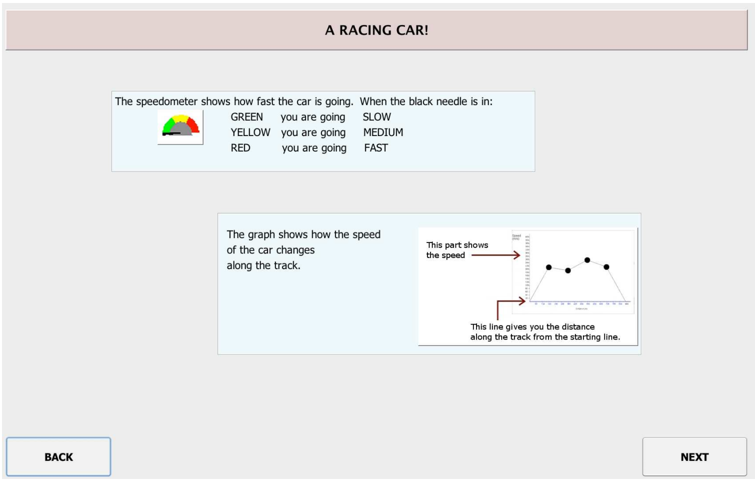


(c) Instructions: Screen 3/3

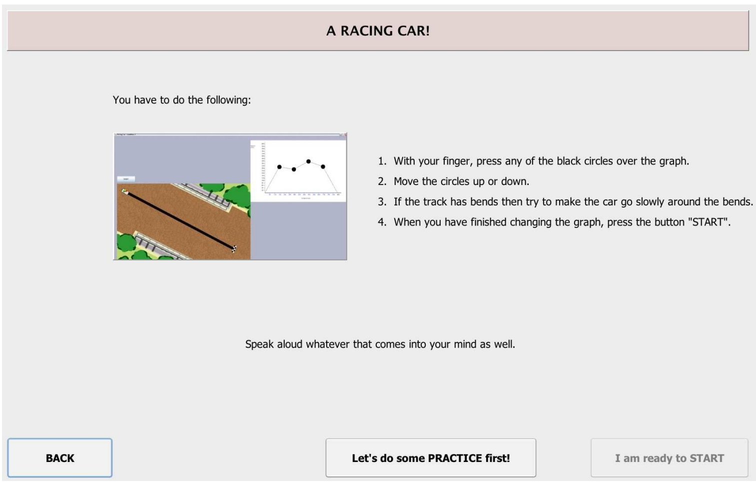
Figure E.2: Screenshots of the Racing Car activity: active_race_car condition.



(a) Instructions: Screen 1/3



(b) Instructions: Screen 2/3



(c) Instructions: Screen 3/3

Figure E.3: Screenshots of the Racing Car activity: active_change_graph condition.

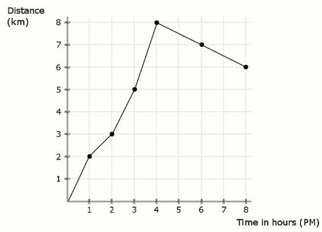
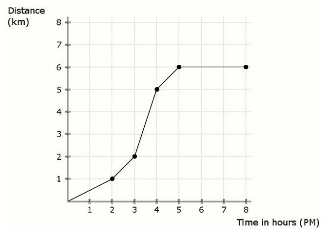
	Battery A	Battery B
	Instructions. Peter is a kid who lives in a very very flat town. He likes to walk and he decided to go for a walk to town. The next screens show a distance-time graph about Peter's trip. Read the questions and answer them using the graph. You can take as much time as you need.	
Line Graph		
Question 1: Reading x-axis	What distance did Peter walk by 3 PM? • 2 km • 3 km • 4 km • 5 km • other answer	What distance did Peter walk by 2 PM? • 1 km • 2 km • 3 km • 4 km • other answer
Question 2: Reading y-axis	How long did it take Peter to walk the first 2 km? • 1 hour • 2 hours • 3 hours • 4 hours • other answer	How long did it take Peter to walk the first 5 km? • 3 hours • 4 hours • 5 hours • 6 hours • other answer
Question 3: Operating two values	How far did Peter walk between 2 PM and 4 PM? • 2 km • 4 km • 5 km • 8 km • other answer	How far did Peter walk between 3 PM and 5 PM? • 2 km • 3 km • 4 km • 5 km • other answer
Question 4: Interpreting	Which of the following sentences describe what happened in Peter's trip? • He walks fast in the first 4 hours and then he walks slow. • He walked slow in the first 4 hours and then he went fast. • He walked fast and then he went down the other side. • He walks up a hill and then goes down the other side. • Other answer.	Which of the following sentences describe what happened in Peter's trip? • He walked slow, a bit fast and then he did not walk. • He walked fast, a bit slow and then really slow. • He went fast, faster and then straight. • He went up a very steep hill and then straight. • Other answer.

Table E.1: Battery 1: Graph Comprehension Questionnaire.

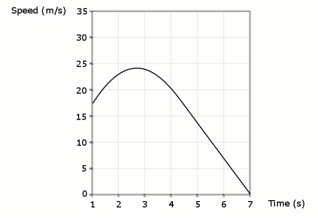
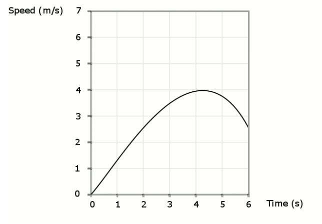
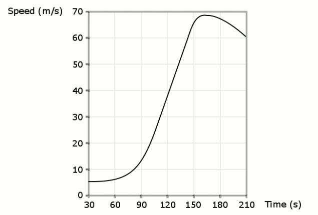
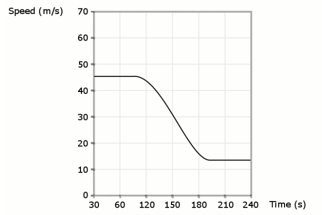
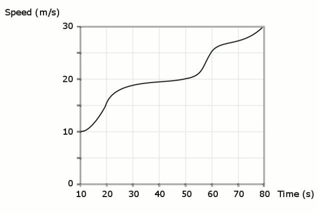
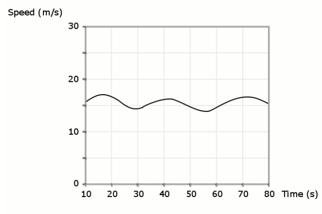
	Battery A		Battery B	
	Instructions. Anne likes to ride her bike around town. She went for a ride 3 times this week. In the next screens you will see speed-time graphs about her different trips. Look at them and select the option that you think shows best what is happening. You can take as much time as you need.			
Question 1		Which best describes what's happening? • She cycled up a hill and cycled all the way down. • She cycled up a hill, went down and then back to the beginning. • She cycled slowly to the bottom. • She sped up a bit and then slowed down. • She was going north and then went south-east.		Which best describes what's happening? • She cycled up a hill and cycled down a bit. • She cycled up a hill, went straight down and then back to the beginning. • She cycled quickly to the top. • She sped up and then slowed down a bit. • She was going north-east and then went south.
Question 2		Which best describes what's happening? • She cycled up an steep mountain. • She cycled up a mountain, rode straight down and back. • She kept speeding up until she reached the top. • She went steady, very fast and slowed down a bit. • She rode northwards and then went south.		Which best describes what's happening? • She cycled down a low mountain. • She cycled down a mountain, rode straight down and back. • She kept slowing down until she reached the bottom. • She went steady, slowed down and went steady again. • She rode eastwards and then went south.
Question 3		Which best describes what's happening? • She cycled over a wobbly road. • She cycled over a wobbly road, went down and went back. • She rode fast across a wobbly part. • She keeps speeding up from time to time. • She cycled to the north-east of the town.		Which best describes what's happening? • She cycled over a bumpy road. • She cycled over a bumpy road, went down and went back. • She rode steady across a bumpy part. • She keeps speeding up and slowing down. • She cycled to the east of the town.

Table E.2: Battery 2: Graph Interpretation Questionnaire (GIQ).

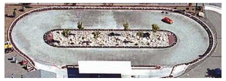
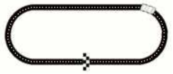
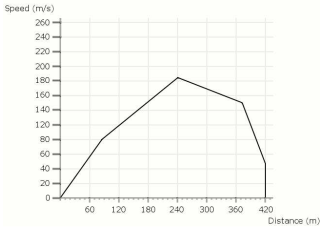
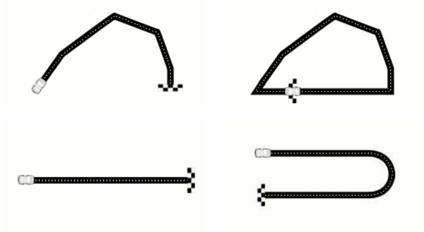
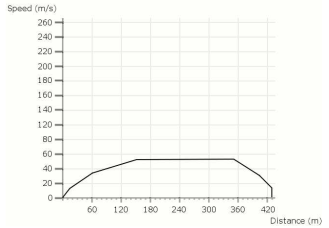
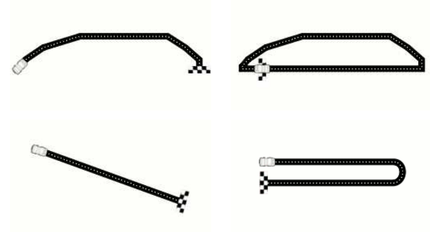
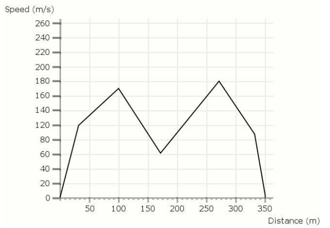
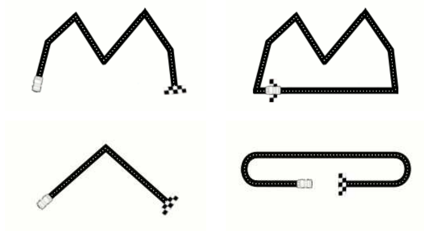
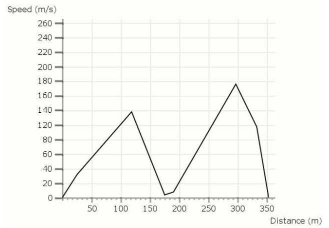
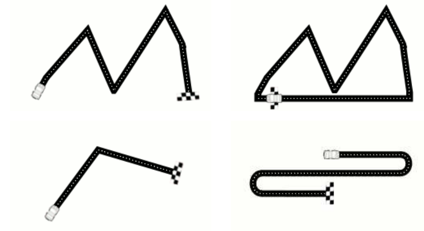
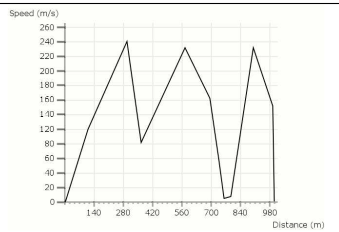
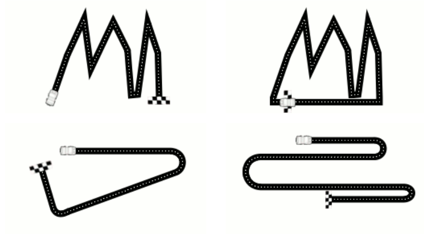
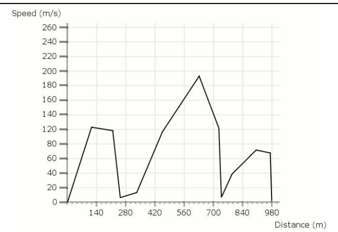
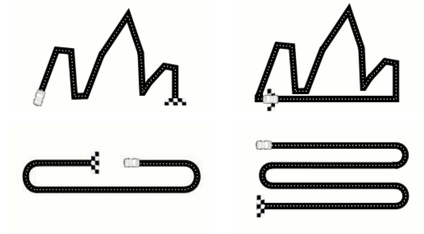
<table>
<tr><td></td><td colspan="2">Battery A</td><td colspan="2">Battery B</td></tr>
<tr><td></td><td colspan="4">Instructions.
James was riding his car around town. He is a smooth driver and he changes his speed depending on the shape of the road. You have to use the following speed-distance graphs to select the road that he might have taken. You can take as much time as you need. Ah! Note that you will see a bird's eye view from above for each road. That is, if the real road looks like this:  then the picture of the road will look like this: </td></tr>
<tr><td>Question 1</td><td></td><td>Which road did James take?
</td><td></td><td>Which road did James take?
</td></tr>
<tr><td>Question 2</td><td></td><td>Which road did James take?
</td><td></td><td>Which road did James take?
</td></tr>
<tr><td>Question 3</td><td></td><td>Which road did James take?
</td><td></td><td>Which road did James take?
</td></tr>
</table>

Table E.3: Battery 3: Graphs and Tracks Questionnaire.

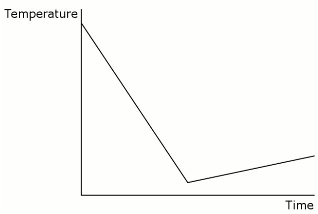
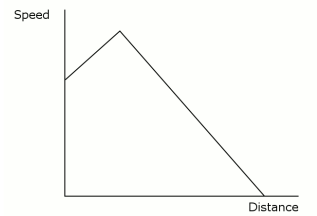
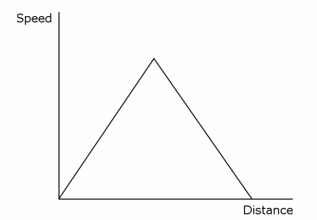
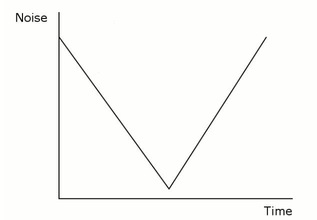
	Battery A	Battery B
	Instructions. What can you tell about the following items?	
Item 1		
Item 2		

Table E.4: Battery 4: Views.

Appendix F

Experiment 2 - Graph Interpretation Questionnaire

Tables F.1, F.2 and F.3 present the distracters used for the Graph Interpretation Questionnaire (GIQ, experiment 2). Details of this battery can be found in Section 5.3.1 and the battery is in Appendix E, Table E.2.

There are two GIQ batteries: Battery A and Battery B; each one has three questions. Each question shows 5 distracters:

1. a correct interpretation
2. pure pictorial interpretation
3. pure pictorial interpretation using x-axis
4. a mixed interpretation
5. a graph-as-map interpretation

The sentences or each distracter are presented in the following tables with their respective graphs. In particular, the sentences are broken down to highlight how they are composed. Mainly, each sentence in Battery A has a parallel item in Battery B. For instance, Table F.1 shows that for question #1 the pure pictorial distracter for Battery A is:

She cycled up a hill and cycled all the way down.

and its parallel item in Battery B is:

She cycled up a hill and cycled down a bit.

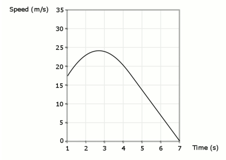
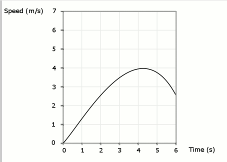
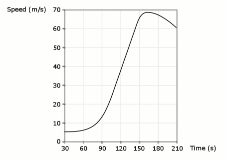
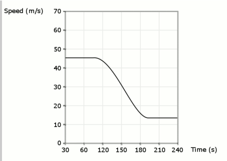
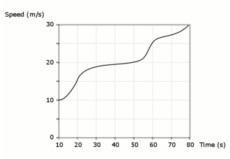
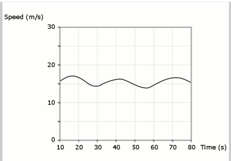
		Which is the best interpretation of this graph?						
			Target-item 1: Picture of the plotted line interpretation			Target-item 2: Picture of the plotted line & axis interpretation		
Question 1	Battery A		She	cycled up a hill	and cycled all the way down.	She	cycled up and down a hill,	went to 7 and came back to 1.
	Battery B		She	cycled up a hill	and cycled down a bit.	She	cycled up and down a hill,	went to 6 and came back to 0.
Question 2	Battery A		She	cycled up	an steep mountain.	She	cycled up a mountain,	rode to 210 and rode back to 30.
	Battery B		She	cycled down	a low mountain.	She	cycled down a mountain,	rode to 240 and rode back to 30.
Question 3	Battery A		She	cycled over	a wobbly road.	She	cycled over a wobbly road,	went down to 60 and back to 10.
	Battery B		She	cycled over	a bumpy road.	She	cycled over a bumpy road,	went down to 80 and back to 10.

Table F.1: Dissection of the distracters for the Graph Interpretation Questionnaire. Table shows the distracters for a pictorial interpretation of the plotted line and the target-items for the pictorial interpretation of the plotted line using axis.

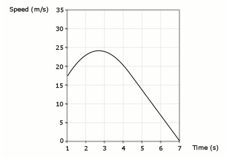
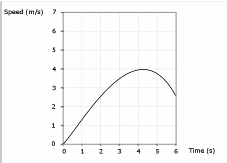
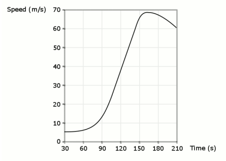
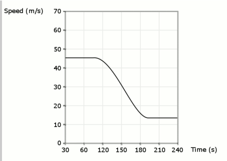
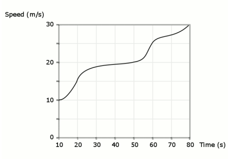
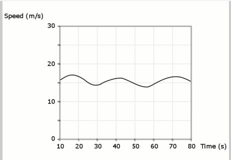
		Which is the best interpretation of this graph?							
			Target-item 3: Mixed interpretation			Target-item 4: Correct interpretation			
Question 1	Battery A		She	cycled slowly	to the bottom.	She	rode	a bit fast	and then slowed down.
	Battery B		She	cycled quickly	to the top.	She	rode	fast	and then slowed down a bit.
Question 2	Battery A		She	kept speeding up	until she reached the top.	She	went steady,	very fast	and slowed down a bit.
	Battery B		She	kept slowing down	until she reached the bottom.	She	went steady,	slowed down	and went steady again.
Question 3	Battery A		She	rode fast	across a wobbly part.	She	keeps	speeding up	from time to time.
	Battery B		She	rode steady	across a bumpy part.	She	keeps	speeding up	and slowing down.

Table F.2: Dissection of the distracters for the Graph Interpretation Questionnaire. Table shows the distracters of the mixed interpretations and the target-items of the correct interpretations.

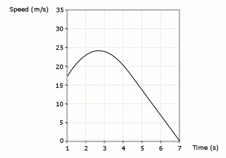
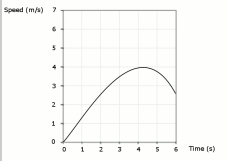
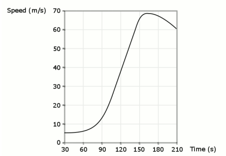
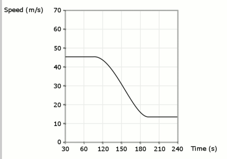
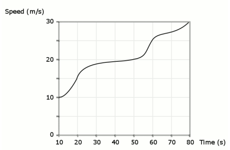
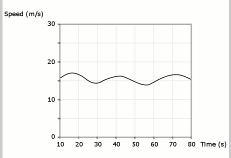
		Which is the best interpretation of this graph?			
			Distracter 5: Graph-as-a-map iInterpretation		
Question 1	Battery A		She	was going north	and then went south-east.
	Battery B		She	was going north-east	and then went south.
Question 2	Battery A		She	rode towards north	and then went south.
	Battery B		She	rode towards east	and then went south.
Question 3	Battery A		She	cycled to the	north-east of the town.
	Battery B		She	cycled to the	east of the town.

Table F.3: Dissection of the target-items for the Graph Interpretation Questionnaire. The table shows the distracted of the graph-as-a-map interpretations.

Appendix G

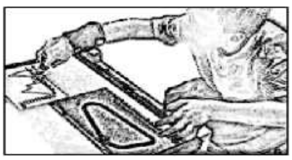
Experiment 2 - Coding Scheme

This appendix contains the description of the coding scheme for the Views battery in experiment 2 (Section 5.3.1). The coding scheme is divided in three parts, named coding dimensions, which are presented after an explanation on how we included gestures in our protocol analysis.

G.1 Gestures

Hand gestures (iconic and deictic, see [51]) will be considered when available. In particular, hand gestures will be used if it is not clear which is the object the student is refering to.

When a deictic (pointing) gesture accompanies the utterance (e.g., *"there"* or *"this"*) then the utterance will be complemented with a description of where the hand is pointing or how it is pointing. For instance, if the student says: *"...when it gets to* ***there****, the car goes about* ***there****..."* then this changes to:



"when it gets to ***there*** **{*right hand index points where the plotted line drops*}***, the car goes about* ***there*** **{*left hand index points to the second bend of the track*}***..."*

Similarly, an iconic gesture will be used to identify which part of the graph is being described when there is not a deictic gesture available. The gesture will be complemented with a description of it.

G.2 Coding Dimension 1: PICTORIAL Conceptions

G.2.1 Purpose

This code has for aim to provide with evidence of the different types of pictorial interpretations. This coding dimension consists of two mutually exclusive codes: `PICTORIAL` and `NON_PICTORIAL`. A `PICTORIAL` code will be assigned to a student's interpretation if and only if there is evidence of at least one pictorial utterance (see Section G.2.3 below). The purpose of having a binary code allows: (1) to clearly separate utterances of pictorial interpretations from the rest and (2) it allows comparing the pictorial interpretations obtained in this battery with the pictorial interpretations from other batteries.

Creating a coding scheme for pictorial interpretations entails describing some of their different forms. The pictorial interpretations observed can roughly be classified into the following categories:

1. **pure pictorial**: (part of the) graph is interpreted as a real world object or as an abstract object.

2. **representational**: (part of the) graph is interpreted as another external representation.

3. **conceptual** or **shaped**: (part of the) graph is interpreted as "an image of an abstract concept" or as a "concept that takes shape" ([51], p. 163).

A more detailed description of those categories is provided in Section G.2.3, but some examples of them might be helpful at this point. Interpreting a graph as a "mountain" or as a "triangle" is considered as a *pure pictorial* interpretation because the graph is mapped with real world object (a mountain) or an abstract object (a triangle). If the student interprets a line graph as coordinates of a map then the utterance is marked as *representational*. A *conceptual* or *shaped* interpretation is found if the student maps the plotted line with a depiction of time (an abstract concept); for example, *"times goes up and down"*.

Note that any mixed conception provided by the student (a partial correct interpretation and a partial pictorial interpretation of the graph) is automatically included in this code. However, mixed conceptions cannot be distinguished with the current binary code. This issue will be addressed when introducing the Coding Dimension 2.

It is important to remark that a `PICTORIAL` code indicates that there is evidence of a pictorial utterance. Therefore, a `NON_PICTORIAL` code only shows that the student has not verbalised a pictorial interpretation. For instance, students could have provided utterances like *"I cannot say"*, *"Nothing"*, *"I don't know"* and those will be marked as `NON_PICTORIAL` because there is no evidence of a pictorial/iconic interpretation.

G.2.2 Procedure for coding

The procedure to code the students' utterances is as described below. Note that (1) in some cases is relevant to observe the students' hand gestures in order to make a more accurate codification and (2) the description of a "`PICTORIAL` interpretation" is in Section G.2.3.

```
FOR EACH (student):
    FOR EACH (item in the battery):
        Transcribe the student's interpretation.
        Complement the meaning of the utterance with the hand gestures.
        IF (there is at least one utterance as described in
           ``PICTORIAL interpretation") THEN
            MARK AS ``PICTORIAL''
        ELSE
            MARK AS ``NON_PICTORIAL"
        ENDIF
    ENDFOR
ENDFOR
```

In addition to the code assigned above, another subcode is given to the utterance if and only if the utterance is marked as `PICTORIAL`. There are three subcodes: `PURE PICTORIAL`, `REPRESENTATIONAL` and `CONCEPTUAL`. These subcodes are assigned according to the categories specified in Section G.2.3. The procedure to allocate those subcodes follows:

```
IF (the utterance was marked as ``PICTORIAL'') THEN
    IF (it was categorized as  ``pure pictorial interpretation") THEN
        MARK AS ``PURE PICTORIAL"
    ELSE
        IF (it was categorized as  ``other representational form
           interpretation") THEN
            MARK AS ``REPRESENTATIONAL"
        ELSE
            MARK AS ``CONCEPTUAL"
        ENDIF
```

```
    ENDIF
ENDIF
```

If the student expresses more than one `PICTORIAL` utterance then only the first pictorial interpretation is sub-coded. This rule is particularly useful when students astray from the topic and it seems more accurate to consider only the first GARM-utterance rather than the subsequent ones.

G.2.3 PICTORIAL interpretation

When the whole/part of a graph is interpreted as an object (real world objects as: mountains, houses), or as an abstract object (e.g., a sound wave, a triangle), or as another external representation (e.g., a map), or it is incorrectly interpreted as a depiction of an abstract concept (e.g., "time goes up and down", the plotted line becomes a "picture" of time), a pictorial interpretation happens. That is, parts of all the graph sections are not given the corresponding abstract meaning, but the interpretation has a literal/iconic/identical[1] correspondence to an (abstract) object, to another representational form or to an abstract concept.

In particular, the idea of "graph-as-a-depiction-of-a-concept" or "graph-as-the-shape-of-a-concept" refers to the mapping of the graph (or parts of the graph) to a pictorial representation of an abstract concept. For example, a student who points at the plotted line and interprets it as *"time goes up and down"*, shows that (1) the student is making explicit the object he/she is talking about: the line that comes "up and down"; and (2) the line is the representation of the abstract concept "time". Thus, the concept takes shape on the graph. This idea is borrowed from the research on gestures from McNeill ([51], p.163) who, in turn, borrows it from Rudolph Arnheim's book *Visual Thinking* [4]. Arnheim asked college students to depict the concepts of past, present and future, showing that a depiction of those concepts can be achieved [4]; but, McNeill suggests that a gesture can be associated with it ([51], p.164) as well.

Furthermore, the graph-as-a-depiction-of-a-concept was found more clearly (but not limited) in cases where the concepts of distance and time are interpreted. Time and distance are concepts obtained from the x-axis (see Appendix E, Table E.4), therefore utterances interpreting time/distance in terms of y-axis variations of the plotted line (which is inappropriate) are identified as wrong/easily. However, codifying concepts that are in

[1] The term "identical" is used in the mathematical sense of an "identity function". That is, the abstraction (e.g., line graph) and the real world object (e.g., a mountain) are treated as the same source/world/(co)domain of the identity function.

the y-axis (speed, noise and temperature) is more challenging. A student uttering *"noise goes down and up"* could have several meanings. For instance, it could mean that the interpretation happens as a pure description of the plotted line; or it could mean that the student is interpreting "goes down" as getting quiet and "goes up" as getting noisier, but the vocabulary is not at hand at the time of the interpretation.

Next, the vocabulary considered for the Coding Dimension 1 is presented. Each part has an abstract description or/and the **key words** (in bold) that were used for coding each student's utterance. Examples for each case are provided too:

PURE PICTORIAL INTERPRETATIONS:

- *"X is **like a** Y"*
  Where X is the whole or part of the graph and Y is the object (concrete or abstract)
  Examples:

  "it is like a slide"

  "it is like a hill"

  "it is like a mountain"

  "it is like a house"

  "it is like a triangle here"

- *"X **is a/an** Y"*
  Where X is the whole or part of the graph and Y is the object (concrete or abstract)
  Examples:

  "it is basically an equilateral triangle."

  "it was a sort of hill"

- *"somebody **goes/gets to**" + (either a description of the plotted line or their fingers point along the plotted line)*
  The plotted line becomes a real object where somebody can go along. For instance, if the graph is interpreted as a hill then somebody can "go down" the hill.
  Example:

  "you sped up and, went down {<u>finger moves along the second part of the plotted line</u>}"

OTHER EXTERNAL REPRESENTATION INTERPRETATION:

- The graph (or parts of it) is interpreted as another representational form.
  For example, the graph line can be confounded with a map; therefore, the abstract meaning given by the student corresponds to the representation that substituted the graph line (the map).
  Example:

  "it is going north east"

DEPICTION OF AN ABSTRACT CONCEPT INTERPRETATION:

- ***"this (is) distance"*** + {*student points to the plotted line*}
  "that (is) distance" + {*student points to the plotted line*}
  "that shows the distance" + {*student points to the plotted line*}
  The plotted line is confounded with the physical extend of the road.
  Examples:

  "that's quite a long distance {student points along plotted line}"

  "that's the distance {student points along the plotted line}"

  "that is showing the distance {hand moves along the first part of the plotted line}"

- ***"this amount of time"*** + {*student points to the plotted line*}
  "for that long" + {*student points to the plotted line*}
  "this (is) time" + {*student points or describes the plotted line*}
  "time is this" + {*student points or describes the plotted line*}
  "time goes" + {*student points or describes the plotted line*}
  The plotted line is confounded and mapped with a period of time.
  Examples:

  "time... it's half down"

  "time went up and down {student points along the plotted line}"

  "in an amount of time {finger moves along the second part of the plotted line}"

- ***"speed is straight"*** or ***"the noise came straight"***
  The concept of speed or noise is confounded with a characteristic of the plotted line.
  Examples:

"speed was going quite fast straight"

"the noise... came straighter"

G.3 Coding Dimension 2: ACCEPTABLE Conceptions

G.3.1 Purpose

This coding dimension identifies acceptable interpretations. An acceptable interpretation is understood as either a correct interpretation using both of the variables involved in the line graph or a "partial" correct interpretation. A "partial" correct interpretation happens when there is evidence of the student interpreting at least one of the variables of the line graph correctly. Thus, there are two codes for this coding dimension: `ACCEPTABLE` and `UNACCEPTABLE`; both are mutually exclusive.

The pupose of this coding dimension is to observe in more detail the varieties of mixed interpretations and it works in conjunction with Coding Dimension 1. Combining both coding dimensions, we can observe 2x2 combinations of students' conceptions:

- `NON_PICTORIAL` + `ACCEPTABLE`
- `NON_PICTORIAL` + `UNACCEPTABLE`
- `PICTORIAL` + `ACCEPTABLE`
- `PICTORIAL` + `UNACCEPTABLE`

Although the aim is to focus on the cases marked as `PICTORIAL`, it is important to highlight that the `PICTORIAL` + `ACCEPTABLE` case is of great interest as, to the author's knowledge, it has not been noticed in any other piece of research before.

Taking into account the `PICTORIAL` + `ACCEPTABLE` codes (*mixed utterances*) and the subcoding from Coding Dimension 1, we can observe the following categories of a mixed conception:

1.1 mixed_pictorial_correct = `pure_pictorial` + `ACCEPTABLE`.

1.2 mixed_representational_correct= `representational` + `ACCEPTABLE`.

1.3 mixed_conceptual_correct = `conceptual` + `ACCEPTABLE`.

In a similar way, there can be combinations of `UNACCEPTABLE` interpretations. However, an `UNACCEPTABLE` interpretation can be marked as such because there was not enough evidence of a correct interpretation (e.g., the student might have not given enough utterances). Thus, it does not provides with such a rich information about the student's mental model as the utterances coded as `PICTORIAL` + `ACCEPTABLE` do.

Observe that a `NON_PICTORIAL` + `ACCEPTABLE` pair assigned to an utterance shows that the student is giving a correct interpretation of the graph.

G.3.2 Procedure for coding

The procedure to code the students' utterances is as described below. Note that (1) in some cases is relevant to observe the students' hand gestures in order to make a more accurate codification (see Gestures, Section G.1), and (2) a description of an "ACCEPTABLE interpretation" is in Section G.3.4.

```
FOR EACH (student):
    FOR EACH (item in the battery):
        Transcribe the student's interpretation.
        Complement the meaning of the utterance with the hand gestures.
        IF (there is at least one utterance that is as described in
        the ``ACCEPTABLE interpretation") THEN
            MARK AS ``ACCEPTABLE"
        ELSE
            MARK AS ``UNACCEPTABLE"
        ENDIF
    ENDFOR
ENDFOR
```

G.3.3 Key-concepts table

In order to identify a correct use of any of the variables involved in the graph, some key words are used as evidence that the student is giving an acceptable interpretation. The key words derive from what are named as **key concepts**, which are listed in Table G.1. Note that the words/phrases are not exhaustive, but they provide an example of the kind of acceptable derivative words.

G.3.4 ACCEPTABLE interpretation

When the student correctly assigns one or more relevant key concepts to the appropriate parts of the graph then the student is giving an acceptable interpretation. Examples of this behaviour are reflected when: the student elaborates an statement accomplishing the

KEY CONCEPT	INSTANCE	COMPARATIVE	SUPERLATIVE	METAPHORICAL DESCRIPTION[a]	AMOUNT/ DEGREE/ DETERMINER of the KEY CONCEPT[b]	OTHER
temperature	hot cold	(it got) hotter warmer colder	the hottest the coldest	It started higher it is high (here) it is lower it is low it goes down it dropped	how hot it is how high the temperature is which is the temperature	
time	early morning noon afternoon late	earlier (in the day) latter (in the day)	earliest latest	it is high it goes past	how much time (at) which time	
noise	loud noisy quiet	louder softer quieter	loudest quietest	it is high it goes down it is (going) lower it lowers down it drops down	how much noise	
speed	quick fast slow	faster slower	fastest slowest	it is high it goes up it comes up it peaked it is highest	how fast	accelerating slowly
distance	far close	further closer	furthest closest	it is high	how far how long	the distance they go

Table G.1: Key Concepts

[a]The metaphorical description must be complemented with a clear reference to the KEY_CONCEPT (temperature/noise/time).

[b]The statement is not a part of a question, but it indicates the amount/degree of the key concept (e.g., how hot, how much time), or it is used to determine the state of the key concept (e.g., which is the temperature). Note that, in this case, “which” can be understood as “whichever”.

function of the graph (e.g., *"how the temperature changes at different times of the day"*); the students specifically describes the variations of one concept in terms of another (e.g., *"the temperature increases in the morning and it decreases in the afternoon"*); the two concepts are interpreted independently (e.g., *"this side the time increases, this side the time increases as well"*); the student is able to interpret only one of the concepts but is unsure about how to interpret the other concept (e.g., *"how the temperature changes, but I am not sure about the time"*).

Although the best scenario is when the student assigns the concepts to the appropriate graph features, it could happen that the student assigns one or more inappropriate concepts to parts of the graph or plotted line (e.g., interpreting *"slowing down"* instead of *"cooling down"* in a temperature vs. time graph). In that case, although the interpretation in another context could be correct, the interpretation for the given context is innapropriate. Therefore, an answer that incorrectly uses one or more relevant abstract concepts from another context will be treated as `UNACCEPTABLE`.

The vocabulary used to code an acceptable answer is given below. Note that, in some cases, observing student's hand gestures is essential.

- ***"what", "when", "how much"***

  The student elaborates an statement accomplishing the function of the graph. The student expreses the extend, amount or degree one concept varies in terms of another. Thus, the two key concepts of the graph are used.

  Examples:

  "what time it was when the temperature was high or low"

  "what time is the noise"

  "how much I have done across time."

- *(interpretation of a key concept) + (its corresponding value in terms of the other key concept)*

  The student describes an state (or a sequence of states) of one of the concepts and the student mentions which is the corresponding value (or sequence of values) for the other concept. Note that the student uses the two key concepts of the graph.

  Examples:

  "$\underbrace{\textit{stops going fast}}_{\textit{value}}\underbrace{\textit{for a while}}_{\textit{state}}$ *then she gets slower in speed"*

$$\text{“}\textit{about}\ \underbrace{\textit{a quarter of the distance}}_{\textit{state}}\ \underbrace{\textit{the speed peaked}}_{\textit{value}}\text{”}$$

- *(interpretation of one key concept only)*

  The student is able to interpret only one of the concepts (using the vocabulary on Table G.1), but either the student (1) is unable to interpret the other, or (2) does not give information about it, or (3) there is an implicit mention of it (e.g., an implicit notion about time). For instance, they might interpret "hotter" and "colder", but they might not be able to interpret how the temperature varies along time. This suggests a difficulty to integrate the two concepts, but it does not explicitly suggests a pictorial type of interpretation. That is, if the student is refering to the plotted line using the concepts associated to one of the variables then it indicates that the student has assigned an abstract meaning to the plotted line instead of a pictorial/iconic interpretation.

  Examples:

  > *"he went quite fast and then slower the speed tells the higher it is the more"* (no other comments about "time")
  >
  > *"the speed bit is saying how fast and distance showing ammm I can't explain it"*
  >
  > *"the noise is dropping down the time, I think, how long it takes to do something"* (interpretation of two concepts independently of one another)

G.4 Coding Dimension 3: Consistency

G.4.1 Purpose

The aim of this coding dimension is to identify the cases where the student's interpretation is inconsistent or unclear. That is, this coding dimension searchers for cases where the student has interpreted the graph in a context that is different from the graph's context. For instance, if the graph is about temperature vs. time and the student mentions anything about "the speed of a car" then the utterance is considered inconsistent.

This coding is of particular help for the cases where the student has given multiple interpretations. Therefore, this coding dimension is a form to eliminate from analysis the cases that are unclear. Note, however, that if a student does not give an utterance then it is considered as if it were consistent (there is not evidence of the student providing an interpretation on one or more different contexts of the graph). For some examples see Section G.4.3.

G.4.2 Procedure for coding

There are three main contexts: temperature, noise and speed (Appendix E, Table E.4). If the student interprets the graph in a context different to the provided context then the utterance is coded as `INCONSISTENT`; otherwise, it is coded as `CONSISTENT`.

```
FOR EACH (student):
    FOR EACH (item in the battery):
        Transcribe the student's interpretation.
        Complement the meaning of the utterance with the hand gestures.
        IF (there is at least one utterance that is as described in
        the ``INCONSISTENT interpretation") THEN
            MARK AS ``INCONSISTENT"
        ELSE
            MARK AS ``CONSISTENT"
        ENDIF
    ENDFOR
ENDFOR
```

G.4.3 INCONSISTENT interpretation

Below, there is a list with the cases considered as inconsistent. This time, there is not a specific vocabulary to follow, but a description of the cases is provided.

- *(switching contexts within the interpretation)*

  The student might have switched the topic of the graph to other topic(s).

  Example:

  > *"the* $\underbrace{\textit{temperature}}_{\textit{context 1}}$ *was quite high down {hand moves along first part of the plotted line} there it is coming down a hill {hand moves along the first part of the plotted line} 'cause it has to probably to* $\underbrace{\textit{speed up a bit,}}_{\textit{context 2}}$ *it's a lot more* $\underbrace{\textit{noisier}}_{\textit{context 3}}$ *about here {hand moves along the second part of the plotted line}"*

- *(appropriate concepts in another context)*

  The student switches the context, but the interpretation in the new context is appropriate. This will not be considered as if the graph is being interpreted as another representation because the line graph is read as a line graph. In this case, note that there is not evidence that the student can make an appropriate interpretation in the context provided.

In the example below, the first utterance is the interpretation for the line graph of temperature vs. time. The second utterance was given for the noise vs. time line graph. See Appendix E, Table E.4.

> *"at higher speed they started almost before they start going slow down and then they sort of gradually get faster"*

> *"because it says 'noise' there... It looks like it's going down quite slowly... right in the middle not so fast"*