Paper prepared for presentation at the 30th International Conference of the System Dynamics Society

July 22-26, 2012, St. Gallen, Switzerland

ASSESSING UNDERSTANDING AND LEARNING ABOUT DYNAMIC SYSTEMS

Birgit Kopainsky, Maria Saldarriaga

System Dynamics Group, Department of Geography, University of Bergen,
Postbox 7800, 5020 Bergen, Norway
birgit.kopainsky@geog.uib.no; maria.saldarriaga@geog.uib.no
phone: +47 47 62 79 86; fax: +47 55 58 30 99

Abstract

One of the main goals of system dynamics models is to improve decision making in dynamic systems. This paper addresses the question of how we can measure what people understand about dynamic systems and what benefit people get from exposure to system dynamics models. For this purpose, we use existing literature about assessing understanding and learning in system dynamics to reflect on outstanding research questions in this area. Learning about dynamic systems requires restructuring of existing knowledge into new knowledge as well as re-use of such new knowledge over time and in different contexts. Existing approaches in system dynamics use elements of dynamic systems to represent knowledge. They thus provide a benchmark for expert knowledge and give indications about the gap between novices and experts. However, they do not provide a theory for further investigating how this gap can be closed. In a second part, we therefore analyze the learning sciences literature for elements that can be useful for the development of a theory about the acquisition, retention, and transfer of knowledge about dynamic systems. We describe first elements of such theory and illustrate how they can help in the design and assessment of dynamic decision-making interventions.

Introduction

Dynamic systems are difficult to understand and manage successfully. This is the case not only in very complex dynamic systems. People also have difficulties making decisions in fairly simple dynamic systems (Brehmer, 1992; Funke, 1991; Jensen, 2005; Moxnes, 1998 & 2004; Rouwette, Größler, & Vennix, 2004; Sterman, 1989a; Sterman, 1989b). Over the years, a variety of strategies have been proposed in system dynamics that aim at improving decision making by supporting learning processes. This paper focuses on one such strategy, the use of gaming-oriented simulations such as simulators and planning games (Maier & Größler, 2000), and within this category specifically on the single-player simulators (as opposed to the multi-player planning games). Simulators link a ready-made simulation model to a human-computer interface and they have various functionalities such as access to additional source materials, explicit presentation of the structure of the underlying simulation model and the progress of time within the simulation etc. (Maier & Größler, 2000).

Simulators are used for teaching purposes and for investigating human decision-making in complex dynamic systems (Größler, 2004). They aim at improving dynamic decision making by providing specific information in the user interface and by guiding learners through specific tasks in the process of working with the simulator (Alessi, 1988; Lane, 1995; Maier & Größler, 2000; Sterman, 1994). This raises the question of how the effectiveness of simulators can be assessed. Assessments or evaluations address two main issues:

- Performance, i.e. the results from decision making. Performance can be measured as the degree to which learners manage to optimize, maximize or minimize a specific measure or how well they reach a specified target (Hsiao & Richardson, 1999).
- Understanding, i.e. the rules that lead to decisions. There are a variety of measures for understanding that range from mean scores in questionnaires to performance in transfer tasks and convergence of mental models between novices and experts (Hsiao & Richardson, 1999).

The majority of evaluations focuses on the first issue and analyzes performance in a dynamic task (for a detailed review see Rouwette, et al., 2004). Performance is a more direct indicator for the quality of dynamic decision making since it is based on learners' explicit decisions and interactions with the simulator. Understanding, on the other hand, is an indicator of the mental models underlying dynamic decision making. It is a measurement of the cognitive, social and motivational learning resources used by the learner during dynamic decision making. Since mental models are not always explicit to the learners themselves, eliciting and measuring understanding is challenging. Another factor that makes assessments of understanding less prominent is the fact that the relationship between understanding and performance is not straightforward and thus neither performance can be predicted by understanding nor can understanding be inferred from performance (e.g., Ajzen, 2002; Berry & Broadbent, 1984; literature reviewed in Doyle, 1997 and Hsiao & Richardson, 1999).

This paper thus contributes to the emerging literature in system dynamics about assessing understanding of dynamic systems (Cavaleri & Sterman, 1997; Doyle, Radzicki, & Trees, 2008; Groesser & Schaffernicht, 2012; Kopainsky, Pirnay-Dummer, & Alessi, 2012; Rouwette, Vennix, & Mullekom, 2002; Schaffernicht & Groesser, 2011). It reviews the documented and measured learning effects in system dynamics based simulators to evaluate *what* learners get from exposure to simulators and *how* this can be measured. Based on this review, we draw conclusions regarding open research questions in this field and use existing literature from learning sciences and education to suggest potential constructs for answering such questions.

Documented and measured learning effects from exposure to simulators

Our study of documented and measured learning effects from exposure to simulators builds on existing reviews that analyze determinants of performance (Rouwette, et al., 2002) and understanding (Hsiao & Richardson, 1999) or list methodological issues concerning the use of simulators in teaching and experimentation (Größler, 2004). In our review, we only include studies that explicitly use *understanding* as a measurement of learning. Studies evaluating *performance* are only considered if they use performance to make inferences about specific changes in understanding. Our review is based on a literature search covering the time period since the last major reviews (i.e., 2000 to 2012) in the proceedings of the System Dynamics Conferences, the System Dynamics Review, as well as Simulation & Gaming. Earlier studies are only included if they provide an explicit basis for subsequent works.

In our review, we focus on *simulator* characteristics and exclude other potential determinants of understanding such as model or player characteristics.[1] First, we focus on the effects of particular instructional strategies in understanding and learning. Then we review the empirical methods used to study such effects and conclude with insights about what we know so far about understanding and learning about dynamic systems.

The reviewed literature on simulators describes a variety of instructional strategies that can be applied when designing simulators. We broadly categorize these strategies into:

- Strategies that are applied when designing the user interface of a simulator, that is, strategies that work on the functionality of simulators. Such strategies include giving feedback following user actions (outcome feedback), explanations of observed behavior (cognitive feedback) and giving hints before user actions (feedforward) (e.g., Alessi, 2000).
- Strategies applied for the design, timing, and sequencing of learning tasks during the interaction of a learner with a simulator. In its simplest form, the interaction of a learner with a simulator consists of reading textual instructions and subsequent decision making trials with the simulator. To this sequence, a variety of learning tasks can be added such as training trials before actual decision making or work with analogies.

Table 1 summarizes the findings of the reviewed studies along these two categories. It describes how modifications in the user interface or changes in the timing and sequence of learning tasks affects understanding. For the purpose of our review, we define *understanding* as the knowledge used by learners to make conclusions and predictions, and we define *learning* as the change in this knowledge.

From Table 1 and the literature summarized therein, the following things can be concluded:

- Most of the modifications in the user interface (e.g., structural transparency and visualization of elements of dynamic complexity) seem to be ineffective in improving learners' understanding of the dynamic system they are interacting with through the simulator. This is particularly the case for visualizations of the elements of dynamic complexity and the provision of entire decision rules.

---

[1] The impact of model characteristics such as length of delays and strength of feedback is one of the main motivations for the use of simulators as they are designed to help learners overcome the difficulties with dynamic decision making caused by these characteristics. Player characteristics need to be controlled for and examined during experimentation.

- The provision of decision cues seems to be more effective than the provision of entire decision rules although the evidence is fairly mixed.
- Cognitive feedback in the form of the provision of a structural explanation of the observed behavior proves to be effective as long as learners get the opportunity to practice dynamic decision making and thus also the use of these decision aids. In general, it seems to be more important whether cognitive feedback is provided or not than in which form this is the case (e.g., causal loop diagrams versus stock and flow diagrams).
- The addition of learning tasks proves to be effective for several designs (e.g., training trials with reduced but increasing complexity; work with cognitive conflict and analogies). All of these designs work with a fairly simple user interface that only provides outcome feedback. The impact on understanding, however, is at least as significant as that originating from additions to the user interface such as revealing the model structure. Adding elements to the user interface does not improve understanding. This is an interesting result in that it indicates that the cognitive load of the user interface can be kept fairly low as long as there are adequate learning tasks involved in the interaction with the simulator.
- Evidence about the effectiveness of simulators in terms of understanding is fairly scarce and scattered across contexts. A few studies systematically build on previous experiments with the same simulator and analyze the benefit of additional elements either in the user interface or in the timing and sequencing of learning tasks. Most studies, however, are based on different simulators and only evaluate their specific effectiveness. This makes their results very difficult to compare with results from other assessments of simulators.
- All studies that find improvements in understanding state that despite the achieved improvements, the distance between novice and expert understanding or the distance to some kind of benchmark remains fairly high.

Table 1: Documented and measured learning effects from the use of simulators

	Design of the user interface (simulator functionality)			
Timing and sequencing of learning tasks	Outcome feedback only	Cognitive feedback: structural transparency	Feedforward	Visualization of elements of dynamic complexity
Textual instructions – DDM trial(s)	Paich & Sterman, 1993 → not effective in improving performance	Größler, Maier, & Milling, 2000, treatment group 2 → no significant increase in understanding; performance significantly lower than in the other treatment groups		Jensen, 2002: → no significant increase in understanding and performance Sawicka & Rydzak, 2007: different representations of delays → no significant increase in understanding and performance Moxnes & Jensen, 2009 → written information about delay does not improve performance
Textual instructions – training trial(s) with full DDM complexity – DDM trial(s)		Langley & Morecroft, 2004 → fastest (and significant) increase in performance with provision of a causal map in the training trials and removal of the map in DDM trials (sustained performance after removal of causal map indicates increase in understanding) Maxwell, 1995 → cognitive feedback less effective in improving understanding than information on decision cues (feedforward) Capelo & Dias, 2009 → strategy map review positively influences mental model similarity Gary & Wood, 2007 → beneficial impact of causal loop diagrams on mental model accuracy	Bois, 2002 → showing which cues to use in the training trials increased performance in the DDM trials and understanding Langley & Morecroft, 2004 → higher performance in initial training trials with the provision of decision cues but no increase in subsequent training and DDM trials Maxwell, 1995: → decision cues more effective in improving understanding than cognitive feedback	Howie, Sy, Ford, & Vicente, 2000: behavior graphs & information arrangement, metaphors, animation to implicitly convey the causal structure → significant increase in understanding; performance significantly better in the second trial

Timing and sequencing of learning tasks	*Design of the user interface (simulator functionality)* Outcome feedback only	Cognitive feedback: structural transparency	Feedforward	Visualization of elements of dynamic complexity
Textual instructions, lecture about model structure – DDM trial(s)	Größler, et al., 2000, treatment group 3 → significant increase in understanding	Größler, et al., 2000, treatment group 1 → significant increase in understanding		
Textual instructions – training trials with reduced but gradually increasing DDM complexity – DDM trial(s)	Kopainsky & Sawicka, 2011 → significant increase in understanding and performance Kopainsky, Alessi, Pedercini, & Davidsen, 2009 → significant increase in understanding and performance	Kopainsky, Alessi, & Pirnay-Dummer, 2011 → no additional increase in understanding and performance with respect to Kopainsky, et al., 2009		
Textual instructions – training trial with full DDM complexity – training trials with reduced but gradually increasing DDM complexity – DDM trial(s)	Yasarcan, 2009 → increase in performance AND understanding (performance not just due to practice)			
Textual instructions – cognitive conflict and analogies – DDM trial	Moxnes & Saysel, 2009; Moxnes & Jensen, 2009 → significant increase in understanding (adequate mental model required for solving the DDM task)			
Textual instructions – DDM trials with full DDM complexity and with live explanations from the experimenter when necessary – DDM trials with structurally similar transfer task	Jensen, 2005 → increase in understanding but not beyond what practice can ultimately achieve			

Methods to document and measure learning effects

Table 2 provides an overview of data collection methods that have so far been applied to measure learning effects from the use of simulators. Some of the studies listed in the table did not test the effectiveness of different interface characteristics or timing and sequence of learning tasks and were thus not part of Table 1. Instead, they measured a change in understanding before and after exposure to a simulator (e.g. Doyle, et al., 2008) or quantified different causal loop diagrams which can be seen as a representation of a mental model of a dynamic system (e.g. Schaffernicht & Groesser, 2011). It is also interesting to highlight that interviews have not been used to assess understanding in studies using simulators.

Independent of the data collection method, the research design for assessing understanding in all the studies included in this review is that of laboratory experiments (for a review of laboratory experiments in the system dynamics field see Arango Aramburo, Castañeda Acevedo, & Olaya Morales, 2012). Surveys or case studies that are applied for assessing the effectiveness of other system dynamics interventions such as group model building (cf., Rouwette, et al., 2002) are absent from the reviewed evaluation of simulators.

Table 2: Data collection methods for measuring learning effects from the use of simulators

Data collection method	**Studies measuring learning effects**	**Aspect of understanding/learning measured by the method**
Think aloud protocols & content analysis	Sawicka & Rydzak, 2007; Jensen, 2002	(The use of think aloud protocols in these studies is mentioned but results are not reported on)
	Jensen, 2005	Concept of equilibrium Instances of static thinking Instances of dynamic control Instances of indirect reasoning Understanding of mutual causation Understanding of the equilibrium situation Use of information Explanation of information
Verbal protocols & content analysis	Kopainsky, et al., 2009; Kopainsky & Sawicka, 2011; Kopainsky, et al., 2011: story questions; coding for understanding of task structure and decision heuristics	Content knowledge/mental models: what learners know Knowledge change/acquisition: change in what learners know Learning leverages: instructional factors affecting learning
	Kopainsky, et al., 2012: story questions; coding for understanding of task structure and decision heuristics; automated analysis for structural and semantic similarity to expert text	Structural and semantic similarity to expert text (mental model similarity)
	Doyle, et al., 2008: open question; coding for translation into a causal scenario diagram and subsequent quantitative analysis of the diagram; empirical application only to test the effectiveness of a simulator-based intervention, not differences in the design of such interventions	Content knowledge/mental models: what learners know Knowledge change/acquisition: change in what learners know Retention: stability of what learners know Knowledge confidence: learners' self-perception of knowledge correctness
Pre-/post-task MC question-	Größler, et al., 2000, learners needed to complete rudimentary CLDs	Content knowledge/mental models: what learners know

naire about model structure	Ritchie-Dunham, 2002: learners needed to indicate the strength of connections between variables	Knowledge change/acquisition: change in what learners know
	Howie, et al., 2000: questions about variables and relationships	
	Kopainsky, et al., 2011: questions about declarative knowledge	
	Gary & Wood, 2007: knowledge test about causal relationships	
Pre-/post-task MC questionnaire about behavior	Größler, et al., 2000: questionnaire about consequences of certain actions in a certain situation (prognoses about future states)	
	Kopainsky, et al., 2011: questions about procedural knowledge	
	Gary & Wood, 2007: questionnaire about consequences of certain actions in a certain situation (prognoses about future states)	
Causal loop diagrams	Groesser & Schaffernicht, 2012: conceptual structure of the content of a mental model of a dynamic system; and corresponding method to measure this structure (Schaffernicht & Groesser, 2011; Schaffernicht, 2006); not tested in the context of simulator-based interventions yet	Similarity to expert model (mental model similarity)
	Capelo & Dias, 2009: revision of strategy maps	
	Mulder, Lazonder, & de Jong, 2011: model structure score that represents the number of correct variables and relations	
Use performance as proxy for understanding	Yasarcan, 2009: regression analysis to rule out practice as main determinant of improvements in performance	(Performance as indirect measure of understanding and learning)
	Moxnes & Saysel, 2009: successful performance only possible with adequate mental model	
	Bakken, 1993; Jensen, 2005: use of transfer tasks	
	Langley & Morecroft, 2004: sustained performance after removing cognitive feedback indicates increased understanding	

Broadly, the methods listed in Table 2 can be classified into two categories according to their definition of the measured aspects of understanding and learning. In the first category, there are studies that are based on the assumption that understanding can be expressed and thus studied in terms of the knowledge used by learners in common forms such as language or writing. This category includes think aloud protocols, verbal protocols and content analysis. These studies focus on learners' verbal and written descriptions of a complex dynamic system and code for elements of dynamic complexity contained in these descriptions such as delays, feedback loops or nonlinear relationships. The second category, on the other hand, includes studies where the focus is on learners' knowledge that can be explicitly represented in terms of existing system dynamics tools such as causal loop diagrams or stocks and flows. Here, learners are required to use specialized terminology and methodology. The assumption is that learners' knowledge of dynamic systems can be represented in terms of system dynamic elements.

Think aloud protocols are evaluated differently in the literature. The use of think aloud protocols originates from the need to generate empirical learning theories for improving performance. This requires the collection and analysis of detailed data on the mental processes that occur during dynamic decision making (Doyle, 1997). Hsiao & Richardson, (1999), in their review, find evidence that under some circumstances, think aloud protocols may help learners perform better in dynamic decision making tasks and thus need to be seen as part of an instructional strategy. Think aloud protocols seem to be a very promising technique for collecting and analyzing data on the mental processes occurring during dynamic decision making. This is emphasized by all studies that apply think aloud protocols and explicitly explore the kind of information that learners look at during their interactions with the simulator (Fu & Gonzalez, 2006; Jensen, 2002; Jensen, 2005).

Conclusions part one: Where do we stand in system dynamics?

Table 3 summarizes the research on the effectiveness of simulators on understanding. The review in Table 1 has shown that evidence on the effectiveness of interface characteristics or learning tasks in simulators is scarce and from different contexts. There are also only few studies that use the same simulator in controlled experimental settings and systematically vary only very few characteristics of the simulator so that the contribution of each element of the user interface and/or simulator-related learning tasks can be assessed. Nevertheless, there seem to be a few findings that have proven to be stable across different contexts.

The table shows that modifications in the design of the user interface to provide cognitive feedback can have a beneficial impact on understanding. Adding learning tasks to the user interaction with the simulator yields yet more benefits in terms of understanding (and performance).

Table 3: Effectiveness of strategies for improving understanding from exposure to simulators

Does not work well	Outcome feedback alone Visualization of elements of dynamic complexity Cognitive feedback – decision rules
Seems to work reasonably well	Cognitive feedback – decision cues Cognitive feedback – structural transparency
Seems to work well	Information strategy using cognitive conflict and analogies Training phase with reduced but gradually increasing complexity

The brief review of methods used for assessing understanding (Table 2) showed that most studies either measure understanding indirectly through performance or they measure understanding in terms of distance to expert understanding (number of correct answers in questionnaires, correct mentions in verbal protocols, quality of causal loop diagrams with respect to correct diagram). Thus, we seem to have adequate methods for determining what expert knowledge of a dynamic system looks like and for measuring how far learners' (novices') knowledge is from this.

All of the reviewed studies that document an improvement in understanding conclude by saying that despite these improvements, understanding is still significantly and considerably below expert understanding. The same is true for studies investigating performance. There is thus still ample room for further improvements. One reason for this is certainly the fact that interactions with simulators are fairly brief teaching interventions and thus cannot be expected to be as effective as other, more long-term interventions such as teaching interventions (e.g., Kunc, 2012; Saldarriaga, 2011; Sterman, 2010; Wheat, 2007) or group model building (for a review of their effectiveness see Rouwette, et al., 2002).

Nevertheless, if we accept the hypothesis that understanding can be further improved also from exposure to simulators, then we need methods that go beyond documenting gaps between novice and expert understanding and that, instead, help analyzing how learners acquire new knowledge during their interactions with simulators and how their existing knowledge hinders or contributes to the acquisition of new knowledge. Especially the studies using think aloud protocols try to identify the kind of information that learners use (or fail to use) when they interact with simulators. This has, however, not happened sufficiently often yet to be able to reach conclusions about how learners acquire knowledge about dynamic systems.

In the remainder of the paper we reflect on how we can become more effective in documenting and measuring learning effects from system dynamics interventions such as simulators. We focus on how to measure how understanding of dynamic systems changes and what inhibits or fosters such change. For this purpose, we study theories and experiences from the learning sciences and reflect on how these can be used in the context of system dynamics interventions.

Insights from learning science research

The connection between system dynamics and the learning sciences is not new (e.g., Bakken, 1993; Larsson, 2009; Mulder, Lazonder, & de Jong, 2009; Saldarriaga, 2011). In this paper, we build on this connection and start collecting insights for the construction of a consistent, comprehensive, and operational evaluation framework for research in learning about dynamic systems. We focus on the study of three phenomena that we believe are essential for learning with system dynamics tools in general and simulators in specific: knowledge acquisition, knowledge transfer and knowledge retention. Learning sciences and education research provide some fundamental insights to study these phenomena. We describe each insight in the subsequent sections. The insights are about conceptual change, units of knowledge and knowledge transfer.

Insight one: Conceptualize learning about dynamic systems as a process of conceptual change

Insight brief: Our first insight is that *learning about dynamic systems should be studied as a conceptual change process and not as the simple addition of new knowledge. In other words, we should look not only at how learners acquire new knowledge of dynamic systems during teaching, but also at how their existing knowledge changes and contributes to or hinders learning.* Our justification for this insight is that unless we understand how learning works in light of learners' existing knowledge of dynamic systems, we will not be able to design appropriate teaching interventions.

Before any formal training, we acquire a significant repertoire of knowledge of causality and behavior that we bring into further learning. In other words, when learning about dynamic systems, we do not come as blank slates. We bring many previously developed ideas that we construct on the basis of our daily life experiences. We do so by attributing causality and resulting behavior based on observations of actions and responses (Kelley, 1973). This intuitive knowledge gives the learner a set of general principles to explain causality and behavior.

Learning research has investigated this intuitive knowledge in a variety of domains. Some of the concepts and phenomena that learning sciences study overlap with the dynamic systems we are interested in (e.g., in mechanics (Brown & Hammer, 2008; Champagne, Klopfer, & Anderson, 1980), and thermodynamics (Lewis, 1996)). Some others are static from a system dynamics perspective (e.g., learners' understanding of the shape of the earth (Vosniadou & Brewer, 1992) and animal classification (Yen, Yao, & Chiu, 2004)), but all of them imply a special kind of learning: a *restructuring* of the learners' existing knowledge, rather than the simple addition of new knowledge. This restructuring of knowledge is commonly referred to as *conceptual change.*

According to Duit and Treagust (2003: 673), conceptual change is necessary "…in such domains where the pre-instructional conceptual structures of the learners have to be fundamentally restructured in order to allow understanding of the intended knowledge…". The system dynamics literature provides ample evidence that dynamic systems fall into such domains (e.g., Cronin, Gonzalez, & Sterman, 2009; Diehl & Sterman, 1995; Forrester, 1992; Moxnes, 2004; Sterman & Booth Sweeney, 2007). Studying learning as a conceptual change process thus implies different research commitments to measure learning:

- For instance, evaluation frameworks cannot only focus on testing whether or not learners have acquired the expert-like knowledge we use as benchmarks. Conceptual change implies change, and change takes time. Thus, evaluating conceptual change requires flexible frameworks that can capture intermediate states between learners' existing knowledge and the expert-like knowledge.
- It is important to identify and describe the ideas that learners have as the learning science literature documents consistently that learners hold ideas that contradict normative scientific principles even after training (Confrey, 1990; Tytler, 2002).
- It is equally important to investigate how these ideas change and become closer to scientific-like ideas (Brown & Clement, 1989; Clark, 2006; Clement, 1993; diSessa, 2007a; Duit, Roth, Komorek, & Wilbers, 2001; Masson & Vázquez-Abad, 2006; Parnafes, 2007).

The subsequent two insights help us describe in more practical terms how to measure all aspects of conceptual change in dynamic systems as well as knowledge transfer and retention.

Insight two: Use proper units of knowledge to study conceptual change

Insight brief: Our second insight is that *to study conceptual change, we need to investigate what specific knowledge changes during learning with system dynamics tools. That is, we need to find the appropriate units to measure knowledge and learning about dynamic systems.* Our justification for this insight is that unless we find appropriate units to measure knowledge and learning, we might be over- or underestimating learners' knowledge of dynamic systems, and we will not be able to investigate how particular teaching interventions work in light of learners' existing knowledge –i.e., what knowledge contributes and what knowledge hinders learning about dynamic systems.

Three particular types of knowledge seem to participate in learners' understanding of dynamic systems. Evaluation frameworks of learning about dynamic systems with system dynamics interventions such as simulators should thus capture and track the development of these types of knowledge from a novice towards an expert-like state: read-outs, causal knowledge, and context-specific declarative knowledge.

Read-outs

An important aspect of understanding of dynamic systems is systems thinking ability (e.g., Booth Sweeney & Sterman, 2000; Kainz & Ossimitz, 2002; Senge, 1990), that is, the ability to see in stocks and flows as well as in feedback loops. Seeing in stocks and flows implies for the learner to focus attention on what is relevant in the particular dynamic system (i.e., the underlying structural elements of the system) and to extract information about these underlying structures.

Learning sciences researchers have developed theoretical constructs to investigate this perceptual aspect of understanding and learning about complex phenomena. The read-outs construct (diSessa & Sherin, 1998) can provide an initial framework to investigate: 1) what elements of dynamic systems learners focus on; 2) what information learners abstract about these elements; and 3) how 1 and 2 change during learning, particularly, though exposure to simulators. By providing answers to these questions, we will be better prepared to explain on an empirical basis what learning to see in stocks and flows and feedback loops implies for the learner. Here, we focus on questions 1 and 2. Question 3 is discussed in our insight 3 (Tracking Changes).

The read-outs construct defines one of the specific types of knowledge that a process of conceptual change of complex scientific concepts must focus on. To reason about complex concepts, a learner must: 1) abstract relevant information (directly accessible or observable) from the system and, 2) connect the abstracted information to the required information (not directly accessed or observed) to make inferences about the system (diSessa & Sherin, 1998). The read-out strategies correspond to the set of tools used by a learner to accomplish the first process. Their name is due to the fact that, in general, the task realized through the perceptual process is to 'read-out' information. The second process will be discussed in the 'causal knowledge' subsection.

To illustrate how the read-out concept is relevant for system dynamics research, take Moxnes' research on renewable resources (Moxnes, 2000). In this study, learners are asked to manage a population of reindeer in order to rebuild an overgrazed level of lichen. The majority of learners abstracts information about the amount of reindeer and assumes a direct effect —a simple cause and effect relationship— of a change in the number of reindeer on the level of lichen. The resulting inference then is incorrect. In contrast, other learners abstract information about the growth and reduction rates of lichen and about the number of reindeer; and assume a stock and flow relationship between lichen, lichen growth and lichen reduction through reindeer. In this case, the resulting inference is correct.

The implications of the read-out construct for an evaluation framework of learning about dynamic systems are that assessments should not only focus on the learners' knowledge about the system's structure and behavior but also on the elements in a simulator (or other teaching material) that learners focus on and what information they abstract about these elements. Low levels of coordination between read-out information and knowledge of the system's structure are the reason why learners fail to provide correct inferences about complex scientific concepts (diSessa & Sherin, 1998). Also, the particular knowledge activated and used by learners in a specific situation depends strongly on the specific 'read-outs' made by them of the situation. It means that the function of the read-out strategies is to activate or point out a specific set of knowledge elements. Thus, we need to understand the relationship between the information we provide to the learner during teaching, the information abstracted by the learner, and the activated knowledge.

Moreover, in the particular case of simulators, read-outs provide a fundamental construct to investigate how learning about dynamic systems is mediated by the specific characteristics of the simulator, especially by the elements of the user interface. Table 4 summarizes how read-outs can be measured.

Table 4: How to measure read-outs

What to measure	1. Aspects (elements, features, properties) of a simulator that become the learner's focus of attention. 2. What information the learner abstracts from these aspects.
How to measure this	Read-outs need to be measured by specific observations (made by researchers) of what a learner says about a system or problem while interacting with a simulator. Video recordings are usually used to collect this sort of data. The data is later analyzed qualitatively

	to develop grounded models (theories) of pairings between specific representations and particular read-outs made by learners (based on what the learner says and does in relation to the representation). Research becomes more scalable quantitatively, after specific grounded models have been proposed and can be further tested with bigger samples.

Causal knowledge

The system's information read-out by the learner needs to be connected to the required information using the learners' causal knowledge. The conclusions made from abstracted information depend on the causal relationships assumed by the learner.

Under the construct of mental models, causal knowledge has been the unit of concern of much work in system dynamics (cf. Table 2). Perhaps the most common research trend is to define mental models from a quantitative perspective, that is, in terms of the number of systems elements they involve and the number of relationships between these elements. Existing mental model frameworks are useful in that they define a benchmark of expert-like knowledge of dynamic systems. They are, however, limited in their effectiveness to describe knowledge change and the knowledge that learners bring before they are exposed to any system dynamics interventions.

The learning science literature provides ample empirical evidence that even after training, learners hold conceptual beliefs about the world that are at odds with scientific concepts and phenomena (Champagne, et al., 1980; Clement, 1982; Driver & Easley, 1978; Johnstone, Macdonald, & Webb, 1977; Trowbridge & McDermott, 1980; Viennot, 1979). By studying these beliefs, it is not only possible to describe learners' knowledge in great detail, but also to identify the particular units of knowledge that are easier or more difficult to acquire or change, or that hinder or support further learning. Moreover, the focus on conceptual beliefs as units of analysis of learners' knowledge has given learning research the possibility of tracking the development of this knowledge during problem solving (McDermott, 1997; Sherin, 2001, 2006), conceptual change(Parnafes, 2005, 2010; Roschelle, 1991), and knowledge transfer (Brown & Clement, 1989; Clement, 1993; Duit, et al., 2001; cf. insight three).

There are different levels of aggregation at which these units are studied (e.g., facets, p-prims, ontologies, etc.). However, most research in this area shares a common concern: to identify units of analysis that properly map learners' ways of thinking (e.g., Clark, 2003; diSessa, Gillespie, & Esterly, 2004; Ueno, 1993; Vosniadou, 2002). For this purpose, novices' knowledge is usually studied in experimental conditions that replicate the real conditions in which such knowledge is usually used by the learner. Moreover, the units of analysis of knowledge are usually defined empirically (bottom-up), rather than pre-defined based solely on expert knowledge benchmarks.

The methods used to measure causal knowledge depend on the nature of the research program. Usually, qualitative research (using methods such as clinical interviews, video recordings of learners' interactions with tools) is conducted to develop grounded theories/models of learners' knowledge and/or learning about a particular domain or range of phenomena (e.g., Roschelle's (1991) micro-analysis of learners' understanding of the concepts of velocity and acceleration). When a theory/model is in place, more scalable quantitative research can be conducted to further test the model (e.g., Hestenes', Wells' and Swackhamer's (1992) Force Concept Inventory).

Qualitative studies of causal knowledge focus on identifying the diverse cause-effect relationships assumed by the learner. However, the "size" of these relationships has also become an important issue. It is assumed, in conceptual change research, that novices' knowledge becomes more comprehensive (complete) and coherent as learners move towards expertise (diSessa, et al., 2004; Vosniadou, Vamvakoussi, & Skopeliti, 2008). And therefore, determining how fragmented or coherent novices' knowledge is, is important if one wants to track how a particular teaching strategy affects such knowledge. In other words, knowledge coherence is an important variable for tracking learning processes.

A way of looking at the issue of coherence is to investigate not only which cause-effect relationships are used by the learner but also, which and to what extent these relationships are assembled together by the learner to form more comprehensive theories. For instance, a learner's knowledge of a particular dynamic system could be constituted by multiple cause-effect relationships assumed by the learner. However, it might happen that these cause-effect assumptions do not constitute a theory of the system for the learner, but only a group of assumptions that he brings together in the spot -even perhaps in unsystematic/random ways (sometimes a particular theory of the system is assembled and sometimes another one is put together). For instance, Saldarriaga (2011) shows how seventh grade students assemble diverse units of causal knowledge to construct an understanding of velocity as a stock –and how this changes during learning. Also, diSessa (2009) shows how learners construct an understanding of Newton's thermal law as a composition of diverse units of causal knowledge.

Another important aspect of measuring causal knowledge has to do with the degree of unconsciousness of such knowledge. Causal knowledge may be rather unconscious to learners: learners know what predictions they make, but they may not be conscious about their underlying assumptions to make such predictions. Research methodologies such as traditional testing are thus not always sufficiently sophisticated to uncover what kind of causal knowledge a learner is using to make a given prediction.

In system dynamics research, the learners' initial knowledge is usually measured under teaching conditions –the same conditions meant to modify learners' initial knowledge. Moreover, learners' knowledge is described in terms of benchmark knowledge –the same knowledge learners are meant to acquire during teaching, e.g., through the use of simulators. In order to observe the actual effect of simulators on learners' understanding of dynamic systems, it seems important that novice knowledge of dynamic system should be investigated using experimental conditions that simulate the real conditions in which learners usually use the knowledge in question. Learners should also be allowed to use all knowledge that they believe is relevant, that is, experimental conditions should not be constrained to activate the knowledge captured by expert benchmarks. Table 5 summarizes how causal knowledge can be measured.

Table 5: How to measure causal knowledge

What to measure	Cause-effect relationships assumed/used by learners when asked to explain or predict a particular situation (system, problem, etc.).
How to measure this	• Methodologies such as clinical interviews and video recording to elicit and collect data of learners' knowledge (Clement, 2000; diSessa, 2007b; Ginsburg, 1997). • Analysis of this data in fine-grained fashion (in short episodes of learning) to identify and track categories/units of knowledge (Parnafes et al., 2008) and to develop theories/models of such knowledge (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; diSessa & Cobb, 2004).

Context-specific declarative knowledge

Context-specific declarative knowledge is when "we *know that xxx* about something" (Novak, 2002). Forms of context-specific declarative knowledge are: equations, narratives that learners memorize from experiences (e.g., "CO_2 accumulation has been increasing in the past years", "human activities increase CO_2") or concept definitions (e.g., radiative forcing).

In system dynamics research we study learners' knowledge of causality but do not necessarily consider the context-specific declarative knowledge needed by the learner to accomplish the dynamic systems task given to them. For instance, Sterman and Booth Sweeney (2007) test learners' understanding of CO_2 accumulation and global mean temperature. They find that, most learners have an incorrect understanding of temperature because they assume an incorrect causal relationship between CO_2 and global mean temperature (pattern matching). However, this lack of understanding might indeed be the result of a lack of context-specific declarative knowledge. Sterman and Sweeney's learners perhaps did not know that "increases in the concentration of greenhouse gases reduce the efficiency with which the Earth's surface radiates energy to space" or that "radiative forcing is the measure of the influence a factor has in altering the balance incoming and outgoing energy in the Erath-atmosphere system" (Sterman & Booth Sweeney, 2007: 220). Although the learners in Sterman and Sweeney's study were provided with all this information (including written descriptions of cause-effect relationships on the system), further testing didn't evaluate whether learners indeed assimilated this information and whether they were actively using it to construct a prediction of the temperature behavior. This information constitutes pieces of context-specific declarative knowledge essential to understanding the causal relationship between CO_2 accumulation and temperature.

A typical example from learning sciences illustrates the relationship between context-specific declarative knowledge and causal relations: Newton's laws of motion. Research has shown that even though learners can declare Newton's laws in equations (i.e., they have context-specific declarative knowledge of these laws and principles), they lack a conceptual understanding of the causal relationships involved in such laws (Peters, 1982).

Including context-specific declarative knowledge in an evaluation framework of dynamic systems knowledge has several benefits:

- It allows having a more complete model of learners' knowledge of dynamic systems.
- It allows designing interventions that address learners' particular difficulties with dynamic systems. Learners may have appropriate causal knowledge but they may lack context-specific declarative knowledge of the particular system that would be necessary to apply this causal knowledge.

Table 6: How to measure context-specific declarative knowledge

What to measure	Factual descriptions used by the learner about the system in question
How to measure this	Context-specific declarative knowledge can be measured in the same ways than causal knowledge. However, context-specific declarative knowledge may be more easily accessible to the researcher (from learners' explanations) since context-specific declarative knowledge is explicit to the learner: the learner knows something and can declare it.

Insight three: Use the units of knowledge to track changes over time and across contexts

Our third and last insight follows from the previous insight: *once we have well-defined units of knowledge of dynamic systems, these units can be tracked to investigate the effects of specific teaching strategies on learners' knowledge change. Moreover, we can track the re-use of particular units of knowledge over time and across contexts to investigate knowledge retention and transfer.* Here, retention is defined as the re-use of a particular unit of knowledge after it was initially learned and transfer is defined as the re-use of a particular unit of knowledge in a context different from which it was initially learned.

This approach allows us to investigate which units of knowledge are more difficult or easier to learn and transfer, and how learners assemble more comprehensive theories out of isolated units of knowledge in the process of learning.

Examples for the insights from learning science research

In this section we describe two examples that illustrate the empirical implications of the insights from learning science research. The first example comes from the learning sciences literature while the second example corresponds to work from the system dynamics community.

Example 1: Measuring read-outs and causal knowledge

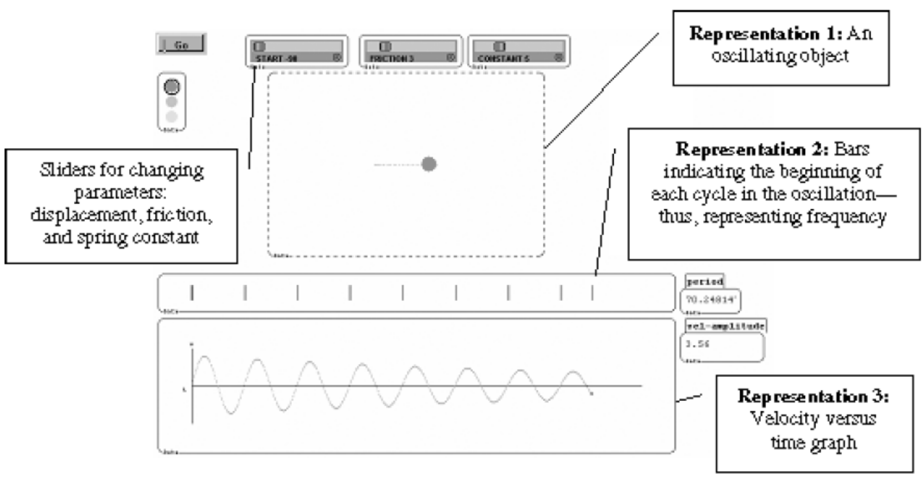
The first example, from Parnafes (2007), illustrates the research methods used to measure read-outs and causal knowledge and to track their development throughout a teaching intervention. In this study, a simulator with a series of *computational representations* is used to support conceptual change of a *dynamic* phenomenon: natural harmonic oscillations. Parnafes investigates how novices' understanding of the concepts of frequency and velocity in harmonic oscillations develop towards scientific knowledge.

Measuring read-outs

What to measure

Parnafes analyses which representation, of the ones available to the learner in the simulation (Figure 1), the learner focuses on, and how the learner interprets such representations. The representations are: an animation of an oscillating object, a bar depicting the periods of the oscillating object, and a time graph of the object's velocity.

Figure 1: Snapshot of harmonic oscillation simulation (source: Parnafes 2007: 422)



How to measure read-outs

Learners are asked to work in pairs. Both, the conversations between the learners and their interactions with the representations in the computer are recorded. The videos are later analyzed to look at what the leaners say and do in relation with the representations. Here, the researcher examines the relation between the use of particular representations in the computer and the change in learners' knowledge. According to Parnafes, her role as a researcher "…was that of a participatory observer. Interventions were made occasionally to clarify meanings or to cue learners to consider overlooked issues"(Parnafes 2007: 421).

Measuring causal knowledge

What to measure

What knowledge learners rely on to make inferences about the oscillatory behavior of objects.

How to measure causal knowledge

Parnafes allows learners to explore physical oscillations with real objects that learners encounter in their daily lives, such as springs and pendulums. Learners interact with these objects in pairs and are asked to talk aloud about their thoughts. They are asked to find and discuss similarities and differences between the different objects. And when they begin discussing concepts such as "fast" or "slow", the researcher asks them to explain this better to focus the discussion around these concepts and to elicit learners' assumptions about frequency and velocity. This stage of the study is done *before* the learners have any interaction with the computational representations designed by the researcher (Figure 1). Once the learners interact with the computational representations, the study focuses on how the knowledge previously used by learners changes. To explore learners' knowledge during their interaction with the representations, they are given tasks for controlling the oscillating object in the simulation (e.g., making the object slow down) and asked to explain the reasons for their actions.

Results from measuring read-outs and causal knowledge

By investigating read-outs and causal knowledge, Parnafes could not only identify the kind of causal knowledge used by learners to reason about velocity and frequency in harmonic oscillatory phenomena, but also how particular features of the computer representations contributed to changing this causal knowledge.

Parnafes found that learners do not differentiate between the concepts of velocity and frequency in harmonic oscillatory phenomena. Instead, learners' understanding of these two concepts is organized around the single lay idea of "fastness", understood as "more X in a unit of time." That is, learners refer to an oscillating object as being fast whether it covers "more distance per unit of time" (velocity) or whether it realizes "more vibrations per unit of time" (frequency). Parnafes identified the assumption of "more X in a unit of time" as being consistent with units of naïve causal knowledge described in earlier research as phenomenological primitives (p-prims; diSessa, 1993). P-prims are basic bits of knowledge that learners assimilate through common experience with the physical world. They are simple, they are not complete model descriptions; they are minimum representations of a given situation. These knowledge bits appear useful to the learner but they become challenging for advancing learning because they are strongly engrained and resistant to change.

Consistently with the previous findings, the focus of Parnafes' teaching study was to investigate how the computer representations contribute to restructuring learners' naïve idea of fastness towards a differentiated and more scientific idea of velocity and frequency. Parnafes found that before the intervention, when interacting with real oscillating objects, learners' abstraction of information about velocity and frequency was opportunistic, that is, they attended to the most available information in front of them (e.g., change in distance in the case of pendulums, and beats of vibrations in the case of springy roads). Parnafes then hypothesized that a change in learners' strategies to focus attention was necessary to change their naïve idea of fastness. Using the construct of read-outs, Parnafes could describe entire episodes of learning where particular learners' read-outs of the computational representations relate to specific changes in their evolution of the idea of fastness. This points to what works and for what it works. For instance, by looking at both representations 2 (frequency bars) and 3 (velocity time graph) in the simulator (Figure 1), a pair of learners (Sue and Robin), began realizing that moving “slower” had two different effects on the object’s behavior: the sine wave in the time graph gets squished and the bars get much further apart. Later on, by controlling the setting for displacement (smaller or higher), Sue and Robin further differentiated between velocity and frequency by realizing that the distance between the bars could stay the same (frequency), even if the peaks of the time graph changed (velocity). Thus, Sue and Robin decoupled the concepts of frequency and velocity by changing their strategies used to abstract available information. From the initial opportunistic strategies, these two particular representations helped Sue and Robin develop more “clear and stable perceptual foci that allowed them to detect the patterns in the simulation…”(Parnafes, 2007: 436).

Example 2: Measuring and tracking causal knowledge during conceptual change and transfer

The second example from Saldarriaga (Saldarriaga, 2011) illustrates the research methods used to measure and track causal knowledge of a dynamic system throughout a teaching intervention using analogies. In this study, an interactive simulation of a water tank analogy is used to support conceptual change of basic dynamics of motion. Saldarriaga investigates how novices’ understanding of Newton’s first and second laws develop towards scientific knowledge conceptualized in stock and flow terms. Since an analogy is used as a teaching strategy, this work requires studying not only conceptual change processes, but also investigating how learners’ transfer restructured knowledge from the water tank to the motion system. In other words, two simultaneous learning phenomena are at play: conceptual change and knowledge transfer. In what follows we describe how Saldarriaga measures causal knowledge, its change and its transfer.

Measuring causal knowledge

What to measure

- What knowledge learners rely on to make inferences about basic motion phenomena (e.g., a toy car moving along a flat surface).
- How this knowledge changes during the tank analogy teaching intervention.
- What knowledge learners transfer from the tank to the motion system
- How learners’ existing knowledge hinders or supports this transfer.

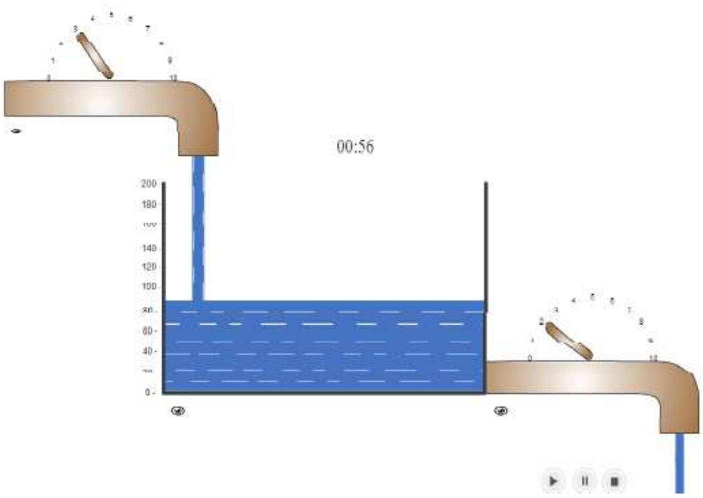
How to measure causal knowledge

Saldarriaga uses a conceptual test, followed by individual clinical interviews to measure learners’ existing causal knowledge before any tank analogy intervention. The test is based on a recognized, thoroughly tested questionnaire called Force Concept Inventory (FCI; Hestenes, et al., 1992). The FCI was designed to test whether learners exhibit any of a full range of naïve conceptions of motion previously identified in extensive research. The follow-up interview follows techniques from cognitive psychology (Clement, 2000; diSessa, 2007b; Ginsburg, 1997) and its purpose is to elicit deeper explanations of why the learner chooses a particular answer in the test. This research strategy has two properties:

- First, it allows studying learners' existing knowledge in a rather ecological manner: using phenomena which learners are more likely to encounter in their daily lives –and which will therefore activate the knowledge that learners commonly use to reason about these phenomena.
- Second, the interview offers a closer look at causal knowledge.

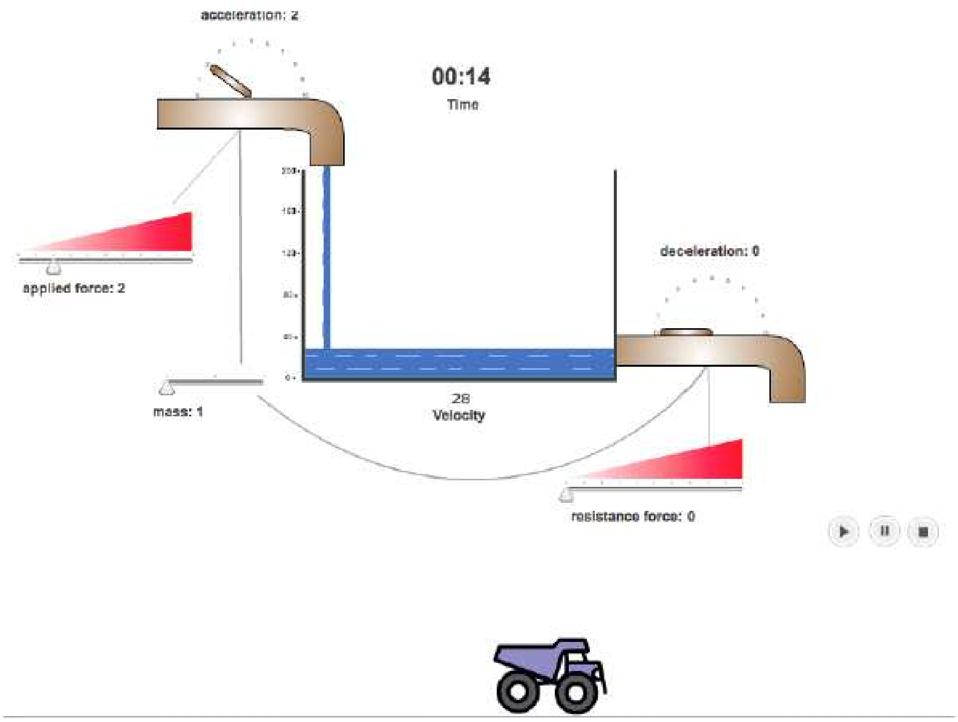
As discussed previously, causal knowledge may be rather unconscious since it is usually "obvious" to the knower. Therefore, measuring causal knowledge may require the researcher to ask learners to "please explain their choices better" several times.

Figure 2: Snapshot of water tank analogy simulator (source: Saldarriaga 2011: 87).



As in Parnafes' study, Saldarriaga has learners working in pairs with a water tank analogy simulator (see Figure 2) to encourage learners to externalize (talk aloud about) their thoughts. Learners are given tasks to control the behavior of the tank system. In a second stage of the intervention, learners are exposed to a modified version of the tank system which shows how the analogy applies to the motion of a toy car that is pushed along a flat surface (Figure 3). In this case, learners are also given tasks to control the toy car's motion. Individual clinical interviews are conducted also after the intervention. The entire intervention is recorded to further support the analysis of the interviews data.

Figure 3: Snapshot of motion simulator (source: Saldarriaga 2011: 99).



Results from measuring and tracking causal knowledge

By using a diverse set of methodologies, Saldarriaga could not only observe the causal knowledge used by learners to reason about basic motion dynamics, but also track the change of this knowledge during the tank analogy intervention.

The knowledge observed had the properties of the same units of naïve causal knowledge observed by Parnafes: phenomenological primitives (p-prims). In contrast to Parnafes however, Saldarriaga tracked not only one p-prim, but 6 different p-prims. By analyzing in detail the episodes of learning in the data, Saldarriaga could describe how each of these p-prims intervened (hindered or supported) in learners' construction of an understanding of the car's motion as a stock and flow system.

For instance, Saldarriaga observed that learners' understanding of *dynamic equilibrium* in motion (i.e., a constant stock of velocity) is hindered by several p-prims. Two of which are *force sustains motion* and *overcoming*. *Force sustains motion* is the idea that a continuously applied force is required to maintain the stock of velocity constant; and vice versa: constant velocity is interpreted as the result of a continuously applied force (an idea similar to pattern matching). *Overcoming* is the idea that constant velocity is the result of an applied force continuously winning over (being greater than) a resistance (opposing) force. By tracking these two p-prims in the interview data, Saldarriaga could observe that:

- First, given the context of the study, some p-prims are more commonly activated than others. Here, *overcoming* was more common (6 out of 11 learners) than *force sustains motion* (2 out of 11).
- Second, the results suggest that how common a piece of causal knowledge is, does not necessarily indicate how difficult it is to change. Here, all learners using *overcoming* exhibited a cor-

rect understanding of *dynamic equilibrium* at the end of the intervention. In contrast, one of the learners using *force sustains motion,* exhibited this piece of knowledge even after teaching.

As Parnafes, Saldarriaga's study is full of descriptions from which models/theories should be formulated and further tested. In particular, there is an important lesson from Saldarriaga's study that invites to reflect on the current methods we use to measure novice knowledge of dynamic systems. Readers may ask the question why p-prims were chosen as the units of analysis. After observing p-prims in her data, Saldarriaga tried to identify whether two or more p-prims were systematically used together by one or more learners. This would have been an indication of learners possessing more comprehensive knowledge structures than the simple, individual p-prims. As none of the learners exhibited more comprehensive knowledge structures than the p-prims, Saldarriaga concluded that p-prims were the right units to measure and track novice knowledge.

This highlights the importance of finding the right units to measure novices' knowledge of dynamic systems. For instance, if Saldarriaga had attempted to track learners' "full theory" of stocks and flows, she would have found that none of her learners had this theory before or even after the intervention. However, her learners did develop their understanding of stocks and flows, although in a more fragmented manner. By using p-prims as units of measure, Saldarriaga could track incremental changes in learners' knowledge of stocks and flows towards expertise.

Conclusions part two: Where should we go in system dynamics?

The examples discussed in the previous chapter highlight the contributions that learning sciences research can provide to system dynamics research. The examples give practical indications of the methods and procedures that can be used to measure those aspects of understanding and learning about dynamic systems that the most widely used methods and procedures in system dynamics cannot measure. The methods applied in the learning sciences allow measuring how knowledge about dynamic systems changes in the course of a teaching intervention and they allow identifying the knowledge structures that inhibit or contribute to such changes. Measuring these aspects of understanding and learning in further research will help improving the effectiveness of simulators and other teaching strategies as they can be made more targeted.

Whether read-outs, causal knowledge and context-specific declarative knowledge are in fact the three main types of knowledge that participate in learners' understanding of dynamic systems is ultimately an empirical question. Further research should test this and further refine the repertoire of elements that constitute learners' knowledge of dynamic systems. For instance, further research should study the role of metacognitive knowledge in learning about dynamic systems.

The examples discussed in the previous chapter also suggest a potential for system dynamics to contribute to learning sciences theory. Dynamic systems impose particular challenges for learning and teaching, as knowledge about dynamic systems not only encompasses knowledge about structure but also about behavior. Studying knowledge about structure as well as behavior may require different constructs and frameworks from the ones already developed. Although the systems studied in the two examples are dynamic, they are rather simple compared to other systems found in the real world. Studying more complex systems may require the development of frameworks that go beyond what current learning sciences research has achieved.

Finally, a general conclusion for this paper is that to study teaching and learning about dynamic systems, we need to develop constructs that are both theoretically and empirically useful. By working collaboratively with a particular set of constructs, we can begin accumulating more effective and comprehensive evidence of *how* system dynamics tools contribute to learning.

Acknowledgements

Work on this paper was supported by the Norwegian Research Council through the project "Simulation based tools for linking knowledge with action to improve and maintain food security in Africa" (contract numbe 217931/F10).The views and conclusions expressed in this paper are those of the authors alone and do not necessarily reflect the views of the Norwegian Research Council.

References

Ajzen, I. (2002). Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology, 32*(4), 665-683.

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-based Instruction, 15*(2), 40-47.

Alessi, S. M. (2000). Designing educational support in system-dynamics-based interactive learning environments. *Simulation & Gaming, 31*(2), 178-196.

Arango Aramburo, S., Castañeda Acevedo, J. A., & Olaya Morales, Y. (2012). Laboratory experiments in the system dynamics field. *System Dynamics Review, 28*(1), 94-106. doi: 10.1002/sdr.472

Bakken, B. E. (1993). *Learning and transfer of understanding in dynamic decision environments.* PhD Doctoral thesis, MIT, Cambridge MA.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology Section A, 36*(2), 209-231.

Bois, R. (2002). *Decisions within Complex Systems: An Experimental Approach Using the Strategem-2 Computer Game.* PhD thesis, State University of New York at Albany.

Booth Sweeney, L., & Sterman, J. D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review, 16*(4), 249-286.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*, 211-241.

Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: abstract transfer versus explanatory model construction. *Instructional Science, 18*(4), 237-261.

Brown, D. E., & Hammer, D. (2008). Conceptual change in physics *International Handbook of Research on Conceptual Change*: Routledge.

Capelo, C., & Dias, J. F. (2009). A system dynamics-based simulation experiment for testing mental model and performance effects of using the balanced scorecard. *System Dynamics Review, 25*(1), 1-34.

Cavaleri, S. A., & Sterman, J. D. (1997). Towards evaluation of systems thinking interventions: a case study. *System Dynamics Review, 13*(2), 171-186.

Champagne, A. B., Klopfer, L. E., & Anderson, J. H. (1980). Factors influencing the learning of classical mechanics. *American Journal of Physics, 48*(12), 1074-1079.

Clark, D. B. (2003). Analyzing Student Knowledge Integration: Theories or Pieces?

Clark, D. B. (2006). Longitudinal Conceptual Change in Students' Understanding of Thermal Equilibrium: An Examination of the process of Conceptual Restructuring. *Cognition and Instruction, 24*(4), 467-563.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66-71.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *J Res Sci Teach, 30*(10), 1241-1257.

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability *Handbook of research design in mathematics and science education*: Lawrence Erlbaum Associates.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher, 32*(1), 9-13.

Confrey, J. (1990). Chapter 1: A Review of the Research on Student Conceptions in Mathematics, Science, and Programming. *Review of Research in Education, 16*(1), 3-56.

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes, 108*(1), 116-130.

Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes, 62*(2), 198-215.

diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction, 10*(2/3), 105-225.

diSessa, A. A. (2007a). Can students re-invent fundamental scientific principles?: Evaluating the promise of new-media literacies *Children's learning in a digital world*: Wiley-Blackwell.

diSessa, A. A. (2007b). An Interactional Analysis of Clinical Interviewing. *Cognition and Instruction, 25*(4), 523 - 565.

diSessa, A. A. (2009). The Construction of Causal Schemes: Learning Mechanisms at the Knowledge Level: Graduate School of Education, University of California at Berkeley.

diSessa, A. A., & Cobb, P. (2004). Ontological Innovation and the Role of Theory in Design Experiments. *Journal of the Learning Sciences, 13*(1), 77 - 103.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus Fragmentation in the Development of the Concept of Force. *Cognitive Science, 28*, 843-900.

diSessa, A. A., & Sherin, B. L. (1998). What Changes in Conceptual Change. *International Journal of Science Education, 20*(10), 1155-1191.

Doyle, J. K. (1997). The cognitive psychology of systems thinking. *System Dynamics Review, 13*(3), 253-265. doi: 10.1002/(sici)1099-1727(199723)13:3<253::aid-sdr129>3.0.co;2-h

Doyle, J. K., Radzicki, M. J., & Trees, W. S. (2008). Measuring change in mental models of complex dynamic systems. In H. Qudrat-Ullah, J. M. Spector & P. I. Davidsen (Eds.), *Complex Decision Making* (pp. 269-294). Berlin/Heidelberg: Springer.

Driver, R., & Easley, J. (1978). Pupils and Paradigms: a Review of Literature Related to Concept Development in Adolescent Science Students. *Studies in Science Education, 5*(1), 61-84.

Duit, R., Roth, W.-M., Komorek, M., & Wilbers, J. (2001). Fostering conceptual change by analogies—between Scylla and Charybdis. *Learning and Instruction, 11*, 283-303.

Duit, R., & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671-688.

Forrester, J. W. (1992). System Dynamics and Learner-Centered-Learning in Kindergarten through 12th Grade Education (Vol. D-Memo 4337). Boston, MA: MIT.

Fu, W.-T., & Gonzalez, C. (2006). Learning in Dynamic Decision Making: Information Utilization and Future Planning (Vol. Paper 19): Department of Social and Decision Sciences.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. Sternberg & P. Frensch (Eds.), *Complex problem solving: principles and mechanisms* (pp. 185-222). Hillsdale, NJ: Lawrence Erlbaum.

Gary, M. S., & Wood, R. E. (2007). *Testing the effects of a system dynamics decision aid on mental model accuracy and performance on dynamic decision making tasks*. Paper presented at the 25th International Conference of the System Dynamics Society, Boston, MA.

Ginsburg, H. (1997). *Entering the child's mind: The clinical interview in psychological research and practice*: Cambridge University Press.

Groesser, S. N., & Schaffernicht, M. (2012). Mental models of dynamic systems: taking stock and looking ahead. *System Dynamics Review, 28*(1), 46-68. doi: 10.1002/sdr.476

Größler, A. (2004). Don't let history repeat itself - methodological issues concerning the use of simulators in teaching and experimentation. *System Dynamics Review, 20*(3), 263-274.

Größler, A., Maier, F. H., & Milling, P. M. (2000). Enhancing learning capabilities by providing transparency in business simulators. *Simulation & Gaming, 31*(2), 257-278.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141-158.

Howie, E., Sy, S., Ford, L., & Vicente, K. J. (2000). Human – computer interface design can reduce misperceptions of feedback. *System Dynamics Review, 16*(3), 151-171. doi: 10.1002/1099-1727(200023)16:3<151::aid-sdr191>3.0.co;2-0

Hsiao, N., & Richardson, G. P. (1999). *In search of theories of dynamic decision making: A literature review.* Paper presented at the 17th International Conference of the System Dynamics Society, Wellington, New Zealand.

Jensen, E. (2002). *Is explicit information important for performance in dynamic systems?* Paper presented at the 20th International Conference of the System Dynamics Society, Palermo, Italy.

Jensen, E. (2005). Learning and transfer from a simple dynamic system. *Scandinavian Journal of Psychology, 46*(2), 119-131.

Johnstone, A. H., Macdonald, J. J., & Webb, G. (1977). Misconceptions in School Thermodynamics. *Physics Education, 12*, 248-251.

Kainz, D., & Ossimitz, G. (2002). Can Students Learn Stock-Flow Thinking? An Empirical Investigation. *Paper presented at the 20th Internation Conference of the System Dynamics Society*.

Kelley, H. H. (1973). The Process of Causal Attribution. *American Psychologist*, 107-128.

Kopainsky, B., Alessi, S. M., Pedercini, M., & Davidsen, P. I. (2009). *Exploratory strategies for simulation-based learning about national development.* Paper presented at the 27th International Conference of the System Dynamics Society, Albuquerque, NM.

Kopainsky, B., Alessi, S. M., & Pirnay-Dummer, P. (2011). *Providing structural transparency when exploring a model's behavior: Effects on performance and knowledge acquisition.* Paper presented at the 29th International Conference of the System Dynamics Society, Washington, DC.

Kopainsky, B., Pirnay-Dummer, P., & Alessi, S. M. (2012). Automated assessment of learners' understanding in complex dynamic systems. *System Dynamics Review, 28*(2), 131-156. doi: 10.1002/sdr.1467

Kopainsky, B., & Sawicka, A. (2011). Simulator-supported descriptions of complex dynamic problems: Experimental results on task performance and system understanding. *System Dynamics Review, 27*(2), 142-172.

Kunc, M. (2012). Teaching strategic thinking using system dynamics: lessons from a strategic development course. *System Dynamics Review, 28*(1), 28-45. doi: 10.1002/sdr.471

Lane, D. C. (1995). On a Resurgence of Management Simulations and Games. *The Journal of the Operational Research Society, 46*(5), 604-625.

Langley, P. A., & Morecroft, J. D. W. (2004). Performance and learning in a simulation of oil industry dynamics. *European Journal of Operational Research, 155*(3), 715-732.

Larsson, M. (2009). *Learning Systems Thinking. The role of semiotic and cognitive resources* (Vol. 145). Lund: Lund University.

Lewis, E. L. (1996). Conceptual Change Among Middle School Students Studying Elementary Thermodynamics. *Journal of Science Education and Technology, 5*(1), 3-31.

Maier, F. H., & Größler, A. (2000). What are we talking about? - A taxonomy of computer simulations to support learning. *System Dynamics Review, 16*(2), 135-148.

Masson, S., & Vázquez-Abad, J. (2006). Integrating History of Science in Science Education through Historical Microworlds to Promote Conceptual Change. *Journal of Science Education and Technology, 15*, 257-268.

Maxwell, T. A. (1995). *Decisions: Cognitive style, mental models and task performance.* PhD thesis, State University of New York at Albany.

McDermott, L. C. (1997). Student's conceptions and problem solving in mechanics *Connecting research in physics education with teacher education*: International Commission on Physics Education.

Moxnes, E. (1998). Not only the tragedy of the commons: Misperceptions of bioeconomics. *Management Science, 44*(9), 1234-1248.

Moxnes, E. (2000). Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development. *System Dynamics Review, 16*(4), 325-348.

Moxnes, E. (2004). Misperceptions of basic dynamics: the case of renewable resource management. *System Dynamics Review, 20*(2), 139-162.

Moxnes, E., & Jensen, L. (2009). Drunker than intended: Misperceptions and information treatments. *Drug and Alcohol Dependence, 105*(1-2), 63-70.

Moxnes, E., & Saysel, A. K. (2009). Misperceptions of global climate change: information policies. *Climatic Change, 93*(1), 15-37.

Mulder, Y. G., Lazonder, A. W., & de Jong, T. (2009). Finding Out How They Find It Out: An empirical analysis of inquiry learners' need for support. *International Journal of Science Education, 32*(15), 2033-2053. doi: 10.1080/09500690903289993

Mulder, Y. G., Lazonder, A. W., & de Jong, T. (2011). Comparing two types of model progression in an inquiry learning environment with modelling facilities. *Learning and Instruction, 21*(5), 614-624.

Novak, J. D. (2002). Meaningful learning: the essential factor for conceptual change in limited or inappropiate propositional hierarchies leading to empowerment of learners. *Science Education, 86*(4), 548-571.

Paich, M., & Sterman, J. D. (1993). Boom, Bust, and Failures to Learn in Experimental Markets. *Management Science, 39*(12), 1439-1458.

Parnafes, O. (2005). The development of conceptual understanding through the use of computational representations. PhD Thesis.

Parnafes, O. (2007). What does "fast'' mean? Understanding the physical world through computational representations. *Journal of the Learning Sciences, 16*(3), 415-450.

Parnafes, O. (2010). When Simple Harmonic Motion is not That Simple: Managing Epistemological Complexity by Using Computer-based Representations. *Journal of Science Education and Technology*, 1-15.

Parnafes, O., Hammer, D., Louca, L., Sherin, B., Lee, V., Krakowski, M., . . . Edelson, D. (2008). How to study learning processes? reflection on methods for fine-grain data analysis: International Society of the Learning Sciences.

Peters, P. C. (1982). Even honors students have conceptual difficulties with physics. *American Journal of Physics, 50*(6), 501-508.

Ritchie-Dunham, J. L. (2002). *Balanced Scorecards, Mental Models, and Organizational Performance: A Simulation Experiment.* PhD thesis, University of Texas at Austin.

Roschelle, J. (1991). MicroAnalysis of Qualitative Physics: Opening the Black Box.

Rouwette, E. A. J. A., Größler, A., & Vennix, J. A. M. (2004). Exploring influencing factors on rationality: a literature review of dynamic decision-making studies in system dynamics. *Systems Research and Behavioral Science, 21*(4), 351-370.

Rouwette, E. A. J. A., Vennix, J. A. M., & Mullekom, T. v. (2002). Group model building effectiveness: a review of assessment studies. *System Dynamics Review, 18*(1), 5-45. doi: 10.1002/sdr.229

Saldarriaga, M. (2011). *Using a water tank analogy to transform students' intuitive knowledge of dynamic systems. A qualitative study of the case of motion.* PhD PhD dissertation, University of Bergen, Bergen.

Sawicka, A., & Rydzak, F. (2007). *Incorporating delays in the decision-making interface: An experimental study.* Paper presented at the 25th International Conference of the System Dynamics Society, Boston, MA.

Schaffernicht, M. (2006). Detecting and monitoring change in models. *System Dynamics Review, 22*(1), 73-88.

Schaffernicht, M., & Groesser, S. N. (2011). A comprehensive method for comparing mental models of dynamic systems. *European Journal of Operational Research, 210*(1), 57-67.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.

Sherin, B. (2001). How Students Understand Physics Equations. *Cognition and Instruction, 19*(4), 479-541.

Sherin, B. (2006). Common sense clarified: The role of intuitive knowledge in physics problem solving. *J Res Sci Teach, 43*(6), 535-555.

Sterman, J. D. (1989a). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes, 43*(3), 301-335.

Sterman, J. D. (1989b). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. [Article]. *Management Science, 35*(3), 321-339.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review, 10*(2-3), 291-330.

Sterman, J. D. (2010). Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review, 26*(4), 316-334.

Sterman, J. D., & Booth Sweeney, L. (2007). Understanding public complacency about climate change: adults' mental models of climate change violate conservation of matter. *Climatic Change, 80*(3-4), 213-238.

Trowbridge, D. E., & McDermott, L. C. (1980). Investigation of student understanding of the concept of velocity in one dimension. *American Journal of Physics, 48*(12), 1020-1028.

Tytler, R. (2002). Teaching for Understanding in Science: Student Conceptions Research, and Changing Views of Learning. *Australian Science Teachers Journal, 48*(3), 14-21.

Ueno, N. (1993). Reconsidering P-Prims Theory from the Viewpoint of Situated Cognition. *Cognition and Instruction, 10*(2/3), 239-248.

Viennot, L. (1979). Spontaneous Reasoning in Elementary Dynamics. *International Journal of Science Education, 1*(2), 205-221.

Vosniadou, S. (2002). Mental Models in Conceptual Development *Model-based Reasoning: science, technology, values*: Springer.

Vosniadou, S., & Brewer, W. F. (1992). Mental Models of the earth: a study of Conceptual Change in childhood. *Cognitive Psychology, 24*, 535-585.

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The Framework Theory Approach to the Problem of Conceptual Change *International Handbook of Research on Conceptual Change*: Routledge.

Wheat, D. I. (2007). *The Feedback Method: A System Dynamics Approach to Teaching Macroeconomics.* PhD thesis, University of Bergen.

Yasarcan, H. (2009). Improving understanding, learning, and performances of novices in dynamic managerial simulation games. *Complexity, 15*(4), 31-42. doi: 10.1002/cplx.20292

Yen, C.-F., Yao, T.-W., & Chiu, Y.-C. (2004). Alternative Conceptions in Animal Classification Focusing on Amphibians and Reptiles: A Cross-age Study. *International Journal of Science and Mathematics Education, 2*, 159-174.