The Role and Nature of Theory in Design Research

Kees Klaassen and Koos Kortland
Freudenthal Institute for Science and Mathematics Education
Utrecht University
The Netherlands

Abstract

We discuss the theoretical input and output of design research, based on our view of what design research should be about.

Input theory is discussed in connection to three aspects of designing: (1) to explain why one wants an educational process that proceeds in a certain way, (2) to explain why the process can be expected to proceed in this way, and (3) to realize one's educational tenets in such a way that one's learning goals are reached.

The first aspect concerns some purpose or value that lies beyond the educational process itself, and to which the process is expected to contribute if it proceeds as intended. A theory to inform one's value-laden choices, if there is a satisfactory one at all, is not object of empirical research. The second aspect concerns the inner workings of the design. Here common-sense psychology is the appropriate guide. It is employed, but not in any way tested. The third aspect concerns the heart of the design process. Here it is as pointless to ask for theoretical guidelines as it is to ask for a theory to be creative.

Concerning output theory we argue that a tested design, if evaluated as good enough, can count as a theory. We dispel worries that more must be done than 'merely' establishing such narrow topic-specific theories. They are just the right kind of theoretical yield given the purpose of science education research, and leave ample scope for progress and comparison.

In our opinion the core business of science education research ought to be the construction, critical discussion and empirical evaluation of detailed content-specific justifications of teaching-learning sequences. It is only when firmly grounded in detailed accounts of concrete teaching-learning activities that general considerations involving theoretical frameworks can be meaningfully clarified, discussed and compared. In this respect much research in science education has it the wrong way round.

1 Introduction

There is still much unclear or unresolved concerning the role and nature of theory in connection to design research. It is unclear, for example, what the role is of grand theories such as behaviourism, constructivism and cultural-historical theory. DiSessa & Cobb (2004) note that such theories are "simply too high-level to inform the vast majority of consequential decisions in creating good instruction" and that "much that is involved in developing specific designs does not follow from these". Nevertheless, in articles reporting a design experiment one invariably reads the claim that in the design one or another grand theory has been applied, implemented, translated, elaborated, transposed, operationalized, embodied, given content, or other such phrases (Méheut & Psillos, 2004; Barab & Squire, 2004). An example of an unresolved issue concerns the nature of the theoretical yield of design experiments. One kind of output consists of narrow domain-specific theories. Researchers who are sympathetic towards design experiments will find this kind of output worthwhile, though a closer look may reveal that their respective conceptions of the nature of domain-specific theories may be somewhat different. There seems to be almost unanimous agreement that science education research, if it is not to reduce to casuistry, ought to aim at more than 'merely' establishing narrow domain-specific theories. Tiberghien (2000), for example, remarks that the design of teaching situations "for each domain of physics can be an endless task". Similarly, diSessa & Cobb (2004) comment "that design research will not be particularly progressive in the long run if the motivation for conducting experiments is restricted to that of producing domain specific instructional theories". Yet it is pretty much unresolved what the nature of a more 'general' or 'broad' theory should be and in what sense such a theory, if there is one, would encode progress.

Based on a particular view of the nature of a didactical design and of design research (Section 2), and of what is involved in testing a design (Section 4), we discuss the role of theory in producing a design (Section 3) and the nature of the theory that comes out of testing a design (Section 5). We close with some general conclusions (Section 6). Throughout the discussion will be supported by examples from design experiments by us and other researchers.

2 A Didactical Design and its Test

A Didactical Design

By a didactical design of some topic we mean a detailed description and justification of the desired (by the designer) development in what students believe, intend to achieve, are pleased about, and so on, in relation to the topical contents, given such and such learning tasks and when guided by the teacher in this and that way (Kortland & Klaassen, 2010). Concerning the justification of a didactical design we distinguish an internal and an external justification.

The internal justification concerns the inner workings of a design. It is an explanation of why the development in students' content related attitudes can be expected to proceed as described, in terms of the designer's estimate of students' initial beliefs and motives and the designer's hypotheses about the outcomes of the specified subsequent learning tasks when guided by the teacher in a specified way.

The external justification relates to why, in the first place, the designer wants a development in students' attitudes that proceeds in this particular way. Here the designer's values enter, concerning the goals he wants students to reach and the educational tenets underlying the ways in which he wants to make students reach the goals (such as that students should be actively involved, or that they should know all along what they are doing and why). The explicit goals and educational tenets together set the quality standard against which the outcomes are to be measured when the design is put to the test.

Design Research

When a design is put to the test in an actual classroom, its internal justification functions as a theoretical prediction – as a hypothetical teaching-learning trajectory. What actually happens in the test provides the evidence in light of which this justification is to be evaluated. The comparison of the theoretical prediction to what actually happens is not straightforward. What actually happens will have to be interpreted in terms of what, at various stages of the process, students believe, mean by what they say, intend to achieve with what they do, and so on. Here triangulation, in the sense of coordinating the interpretations of various researchers, is useful, if only to avert the danger of seeing what one hopes to see (one's predictions). Proceeding in this way, the designer can make his interpretations as rigorous, systematic and objective as can be.

Design research that is conducted in this way aims to improve on the practical wisdom of experienced teachers, both in being more detailed and specific with respect to expectations beforehand and in being more systematic and impartial in evaluating whether or not these expectations have come true. Furthermore it aims at more than what can be achieved by a pre-test post-test research design. The latter may give an indication *that* a teaching-learning sequence works or fails to work: are the intended learning goals reached, as measured by the progress from pre-test to post-test? Design research, as we envision it, also aims to understand in detail *how* and *why* the teaching-learning sequence works or fails to work: does the teaching-learning process itself proceed as hypothesized? It is precisely this detailed content-specific understanding of the process that promises to offer a worthwhile, evidence-based resource to guide professional practice. We will come back to this claim in Section 5, but before that we already hope to have made the claim plausible by means of the examples we provide.

3 The Role of Theory in Producing a Design

Justification of Topical Goals

Many of the arguments that are typically given to justify value-laden choices we find wanting. We read, for example, that this or that subject or topic should be taught, because it is needed for students to participate in a highly technological democratic society, because it is relevant for students' daily life, because it provide students with a challenging view of science, because it acquaints students with the sorts of professional practices in which they are likely to get involved in their future working life. However, we feel that our physics knowledge hardly contributes to decision making on difficult societal issues, such as whether or not to build new nuclear power plants. We do not believe that the science behind food processing or perception, if we were able to master it at all, would greatly enhance our everyday cooking or observational practice. We admit that students are very keen on devices such as mobile phones and iPads, but we do not believe that they are also very keen on finding out how these devices work. And why already bother students at school with the practices they will be involved in for the rest of their working lives? Let us be clearly understood. We are not arguing to stop teaching science. We are just questioning arguments in favour of it.

Our main message is that everybody, we ourselves included, is in a bad shape when it comes to justifying topical goals. Perhaps it is possible to improve the argumentation.

Perhaps it is possible to devise a theory that could help us any further here – some normative scheme for weighing various subjects and topics according to the contribution they make to some societal ideal. However, all of this is far beyond us. Furthermore, if there was such a theory, we do not see how it could ever be object of empirical research.

When it comes to topical goals our pragmatic attitude is to be pretty liberal. We simply ignore the argumentation in favour of it, and take the adopted goals for granted or see them as expressions of personal preferences. Nevertheless, the topical goals cannot simply be decided on in advance. Whether they can be realised with sufficient quality, as measured by the designer's own standard, will have to be investigated.

Justification of Educational Tenets and the Role of Grand Theories

Also with respect to educational tenets we are not very impressed by the argumentation. In our opinion, derivations of educational tenets from grand theories or orienting frameworks, such as behaviourism, constructivism or cultural-historical theory, are not proper arguments at all. Take for example the educational tenets that science should be taught in such a way that leaners are actively involved, or that they are given ample freedom to make their constructions explicit. Often such tenets are claimed to follow from constructivism. We do not see, however, how they follow from the basic constructivist idea that new knowledge is constructed on the basis of already existing knowledge. If this idea is true, then no matter how a learner is taught, he/she will construct new knowledge on the basis of what he/she already knows. It does not help to claim that 'active involvement' or 'freedom of construction' will lead to improved learning. This does not follow without introducing additional values of what is to count as 'better' learning. Let us be clearly understood. We are not questioning the value of 'active involvement' or 'freedom of construction' in itself. We merely deny that these values follow from one or another grand theory. More generally, learning theories do not entail normative decisions about how to teach (Millar, 1989).

We agree with diDessa & Cobb (2004) that grand theories do not "inform the vast majority of consequential decisions in creating good instruction", and we add that no theory whatsoever does or can do. Designs that are 'based on' the same theoretical starting points turn out to be totally different and, conversely, designs that are 'based on' totally different starting points turn out to be remarkably similar (De Klerk, 1982). In design studies it is often claimed that some specified grand theory has been leading for the design, or that the theory directed, steered, guided or supported the design. But such claims are hardly ever seriously substantiated or continently argued for. That is, one mostly searches in vain for clear answers to such questions as: Which decisions did you by means of which reasoning derive from which theoretical principles? Could the very same decisions not also have been derived from other theoretical principles?

Our view of the role of grand theories is that researchers typically feel indirectly supported by one grand theory or another, in that they would not have done what they did if they had not known that theory. In this sense we see the role of grand theories mainly as one of clarification, by means of which researchers can indicate what their sources of inspiration are. This does not imply that we are as liberal with respect to educational tenets as we are with respect to topical goals. If a grand theory has serious shortcomings – as we think is the case for the grand theories mentioned above – then as an orienting framework it may orient the designer in a completely wrong direction. Take, for example, the tenet to take account of students' ideas. We think things go seriously wrong if this tenet is derived from the idea that students' views of the world are alternative to the scientific view of the world (see Example 4 later on in this paper).

Internal Justification

What is needed to deliver an internal justification of a didactical design simply is the background theory of which everybody knows *that* it needs to be used, and *how* it needs to be used, in order to find out about and influence the mental life of others. Let us call it common-sense psychology. Making explicit the reasons for one's expectations about how the teaching-learning process will proceed, may in itself already be sufficient to bring to light quite a lot of wishful thinking. Here too triangulation, in the sense of discussing one's design with colleagues, will make the expectations more realistic by diminishing cases of tunnel vision.

Below we present an example of an internal justification. The example concerns a series of tasks taken from the first version of a teaching-learning sequence about radioactivity, developed at the end of the 1980's (Klaassen, 1995). Note that this was just a few years after the accident with a nuclear power plant in Chernobyl in 1986. Students at the time were all familiar with the accident and its consequences. In the Netherlands, for example, fresh products such as milk and spinach had to be withdrawn from the market. After the presentation of the series of tasks we will provide some context and indicate the educational work that we intended the series of tasks to deliver, which has to do with our topical goals and educational tenets. Then we present the internal justification of the series of tasks, i.e. an account of why we expected the series of tasks to deliver the intended work, and the role of common-sense psychology in this account.

Example 1: The Chernobyl Accident

The Series of Tasks

In the teaching-learning activity reproduced below the pictures that were originally included are merely described. The term ‘R-sources’ in the text had been introduced in Chapter 1 of the textbook to characterise an object or apparatus in the vicinity of which a Geiger counter starts ticking.

2-3 SOURCES WITH RADIOACTIVE MATERIAL

In the previous section you have noted that R-sources need not always be dangerous. At some distance from a jar with radioactive stones, for example, it no longer is dangerous.

The stones in the jar emit *radioactive radiation*. That is why we call those stones *radioactive*. At some distance one hardly measures the radioactive radiation that is emitted by the stones. The Geiger counter then ticks as fast as when the jar with stones had not been there.

This is the same with sound. If one stands at a far enough distance, one cannot hear (or measure) the sound that is emitted by a radio.

4 Up to how far can the radiation from the jar with radioactive stones still be measured? How did you check that?

[Picture of some nuclear fuel elements, with the caption: "In a nuclear power station there are rooms in which radioactive material is stored. Those rooms have thick concrete walls."]

5 The jar with radioactive stones is an example of a source that contains *radioactive material*.
a Do you know any other sources that contain radioactive material? Which?
b Up to how far do you estimate that the radiation from those sources can still be measured? Why do you think so?

6 In the Netherlands there are two nuclear power stations. In Borsele and in Dodewaard.

[Map of the Netherlands, in which Borsele and Dodewaard are indicated. Pictures of the nuclear power stations in those places.]

Do you think that where you live one can still measure radiation from those nuclear power stations? Why?

[Picture of the nuclear power station in Chernobyl before the accident.]

Unfortunately, accidents happen now and then in nuclear power stations. Perhaps you can still remember the accident with a Ukrainian power station in Chernobyl that happened in 1986. The picture above shows the power station in Chernobyl before the accident. Below after the accident.

[Picture of the nuclear power station in Chernobyl after the accident.]

7 Also in the Netherlands radiation was measured because of the accident in the nuclear power station in Chernobyl. Chernobyl is more than 1500 km away from the Netherlands.
a Could we in the Netherlands measure radiation from the nuclear power station in Chernobyl before the accident had happened? Why?
b How come that after the accident we could measure radiation from it in the Netherlands?

8 The jar with radioactive stones is standing in front of the classroom. You have already learned that at the back of the classroom you won't receive radiation from it.
Think of what would have to happen in order that at the back of the classroom radiation can be received by it.

The Intended Educational Work of the Tasks

Before the teaching-learning sequence was designed, Klaassen (1995, sections 2.3 to 2.5) first did some research on students’ existing knowledge about radioactivity. The findings were that students’ existing knowledge could to a large extent be understood in terms of very basic notions concerning causation. In essence, and in our wordings, an affector harms an object by means of an instrument. In the case at hand, X-ray machines, radioactive waste, irradiated food, Chernobyl, and so on, have the potential to harm something or someone, because in one way or another they can make it happen that something harmful enters the thing or person. This ‘something harmful’ is often called ‘radiation’ or ‘radioactivity’ by students. It functions as the instrument. In the case at hand it is invisible, transportable and penetrating. The Chernobyl accident was an affector, because huge amounts of the instrument were released. According to many students, irradiated food is a potential affector, because it contains the instrument and by eating the food we get the instrument inside. An object or person is affected as long as it contains the instrument. The effects may be reduced by applying a resistance, i.e. something that counteracts the instrument. A resistance such as a lead wall or a special suit prevents that the instrument enters an object or person.

A major aim of the teaching-learning sequence was to make students differentiate the general idea of a harmful instrument into the separate notions of radioactive substance and radiation, in the context of understanding safety measures. Whereas initially students are likely to judge every situation having to do with radioactivity as dangerous, they are to learn that one can easily protect oneself against closed sources by keeping some distance, that objects will not emit radiation after being irradiated, and so on.

Within this overall aim, the work that Tasks 4 to 8 were to deliver is that students would find out something like this: in order that radiation can be measured at a large distance from a radioactive source, it must have been the case that radioactive material has escaped from that source. In other words, that students would come to value the distinction between open and closed sources.

The intended contribution of Task 4 was to make students recall the earlier finding that at some distance from the jar the radiation emitted by the radioactive stones can no longer be measured. The intended contribution of Tasks 5 and 6 was to make students believe, aided by the suggested analogy in the pictures (rooms with thick concrete walls in which radioactive material is stored), that if one stands at a far enough, but still rather limited distance, one also will not measure any radiation emitted by the radioactive material in other sources (Task 5) and in particular nuclear power stations (Task 6). The intended contribution of Task 7a was to make students extend this conclusion to the case of the power station in Chernobyl before the accident (no radiation could be measured from it in the Netherlands), and therefore to see the point of Task 7b. The intended contribution of Task 7b was to elicit familiar knowledge about what happened at and after the accident in Chernobyl, for two purposes. First, in order to explain why due to the accident radiation could be measured in the Netherlands: there was an explosion in the Chernobyl plant; the wind was directed towards Western Europe; it rained in the Netherlands, and so on. Second, in order to make proposals, again aided by the analogy, about what would have to happen in order that at the back of the classroom radiation can be measured: open the jar; put a fan behind the jar to produce a draft toward the back of the classroom, sprinkle some water there, smash the stones to pieces, and so on.

The intended contribution of Task 8 was to provide a context in which students could actually check if their proposals worked (as far as safety allowed, of course). The intended contribution of these checks was to make students use whatever they were to find out (e.g. that simply opening the jar is not sufficient) in order to (re)consider, again guided by the analogy, how it can be that due the Chernobyl accident radiation could be measured in the Netherlands. Finally, to make them eventually conclude that it must be due to the radioactive material that escaped after the accident (and that by wind, rain and so on was brought to the ground in the Netherlands).

Internal Justification of the Tasks

Why did we think that the tasks would deliver the intended educational work? Well, the reasoning involved is so simple and obvious that one is apt to miss it. Consider for example the intended contribution of Tasks 5 and 6: to make students believe, aided by the suggested analogy in the pictures (rooms with thick concrete walls in which radioactive material is stored), that if one stands at a far enough, but still rather limited distance, one also will not measure any radiation emitted by other sources containing radioactive material (Task 5) and in particular nuclear power stations (Task 6). In order to spell out our reasoning, we first of all need to mention some of the beliefs that we assumed students to have at the time of working on the tasks: (1) relevantly similar things behave relevantly similarly under relevantly similar circumstances; (2) the jar with stones is relevantly similar to a nuclear plant in that both are closed containers with radioactive material, and although a nuclear plant contains much more radioactive material than the jar, nevertheless the circumstances are relevantly similar in that the walls of the plant are much thicker than those of the jar; and (3) at some distance from the jar with radioactive stones one hardly measures any radiation from it.

Now, we may of course have been wrong in assuming that this is (part of) what students believed at the time. But the main point to make here, in view of clarifying the internal justification of the tasks, is that from propositions (1) to (3) it follows logically that if one stands at a far enough distance from a nuclear plant one will hardly measure any radiation from it. And this precisely is why we expected Tasks 5 and 6 to deliver the intended work. That is, we expected the tasks to make students believe the latter proposition, simply because it is a logical consequence of what we assumed them to believe. Or perhaps better, if their work on Tasks 5 and 6 would not lead students to believe that proposition, we would not question students’ logical reasoning abilities, but rather question our assumption that students believed all of the propositions that together entail the latter one. It simply is constitutive of the concept of belief, i.e. part of what makes a belief the belief it is, that beliefs are logically consistent with each other.

And therefore it is a basic element of attributing beliefs to students that we are constrained to do so in such a way that they also turn out to believe the logical consequences of the beliefs attributed to them. It is one of the basic elements that we use all day long to find out what others believe. In other words, it is a basic element of what we have called common-sense psychology. What this particular element does, is to demand that beliefs fall into rational patterns, in this case patterns of entailment relations. What the internal justification consists in, in the example just given, is to explicitly and concretely display such a rational pattern.

Apart from displaying entailment relations, much more is involved in internal justification. Part of the internal justification of Task 8, for example, is to explain why students are expected to propose to open the jar. Spelling out the reasoning now boils down to the assumption that at the time students (1) want to make it the case that radiation can be measured at the back of the classroom and (2) believe that opening the jar will produce that state of affairs. From this belief-desire pair it simply follows that there is something attractive about opening the jar. If students did not propose to open the jar, and did not advance any other proposal that according to them will make it the case that radiation can be measured at the back of the classroom, this would count against attributing to them the desire to produce that state of affairs. Again, it is a basic normative element of common-sense psychology to transfer the perceived value of an end to the perceived value of a means to reach that end. Another, trivial, example concerns the explanation of why it can be expected that students come to believe that the Geiger counter is ticking when they hold the counter near the jar with radioactive stones. The reasoning simply is that under the circumstances the Geiger counter *will* be ticking. Again, it is a basic element of common-sense psychology, that plain perceptual beliefs are about what causes them.

In all the cases illustrated above it is common-sense psychological reasoning that allows us to justify the inner workings of the didactical design, simply by displaying the causes of and patterns in what students believe to be the case, want or intend to make the case, and so on. We refer to Davidson (2004, essay 10) for a more precise characterization of the rational patterns demanded by common-sense psychology, and for an informal proof that together they suffice for understanding.

Producing a Design

There is no theory that serves to actually produce a design, just like more generally there is no theory to be creative. The designer will without doubt benefit from a thorough study of the science involved, its relations to technology and society, its philosophical foundations and its historical origins. Furthermore the designer may well be inspired by one or another philosophical, psychological or learning theory. His educational tenets and associated design principles will make him receptive for useful ideas and make him recognize a good idea as such. But the educational tenets and associated design principles do not play a facilitating role in actually generating particular teaching-learning activities. This process, just like any creative process, is a matter of finding local solutions to local problems – and this is what we consider to be the heart of design research. Success or failure may critically depend on details such as the actual wording of tasks (Viennot, 2003). Similar remarks apply to the process of adjusting a design in light of the test results. There just is no way to regularize this process, just like more general there is no way to regularize the process of creating new theories to cope with new data. It requires skill, talent, persistence, and a good deal of luck.

An example of such a process, in this case concerning a teaching-learning activity called ‘The Comet Kirch’, is given below. The activity is part of an introductory mechanics course for students in upper secondary education (Klaassen et al., 2008; Emmett et al., 2009). It must be noted beforehand that an external justification for this course is lacking: Why bother 16-year olds with mechanics? Is mechanics, however fundamental it may be regarded in physics, an essential or even useful part of a preparation for their future life?

Example 2: The Comet Kirch

Educational Tenet: A Problem-Posing Approach

The educational tenet of a ‘Problem-Posing Approach’ (Klaassen, 1995) is that a teaching-learning process should proceed in such a way that all along students know, on content-related grounds, what they are doing and why. This tenet is indirectly supported by Davidson’s philosophy (Davidson, 2006). His theory of interpretation has led us to the general view that, if properly interpreted, students’ existing thought and action in general do not have to be corrected from a scientific point of view, and, moreover, contain useful seeds for an extension towards science. From this point of view, the educational challenge becomes one of making students add substantially to their existing knowledge and skills, even though there is not much wrong with their existing knowledge and skills. Hence the emphasis on so providing students with content-related motives and on so soliciting seeds in their existing ideas, that they are willing and able to extend their knowledge and skills in the intended direction.

One way to provide students' learning of mechanics in advance with a sense of direction is to appropriately tap core causal reasoning, in particular notions such as 'changes are caused' and 'things go on as they are unless interfered with'. Explanation of motion is a special case of causal explanation, in which one is not concerned with changes of state in general, but with changes of state of motion, and where forces are the effectuators of these changes.

In his first law of motion, Newton explicitly states that of its own accord (force-free motion) a body would uniformly move forward in a straight line. Deviations from this motional state are caused by forces that act on the body. This is illustrated by Newton's construction method in Figure 1 (see also Newton, 1999, p. 444). An attractive force directed towards the sun S influences the force-free continuation B–c of the motion A–B of a planet. The actual motion B–C is the superposition of the force-free continuation and the deviation B–V caused by the force. Other assumptions for force-free motion are also possible. Kepler assumed *rest* to be the force-free motional state, at least for celestial bodies. In this case the construction is rather different (see Figure 2). Here a force is required to 'push' the planet from A to B, from B to C, and so on, since of its own accord the planet would remain in one place. So Kepler had to somehow find forces that could 'drag' the planets along their paths.

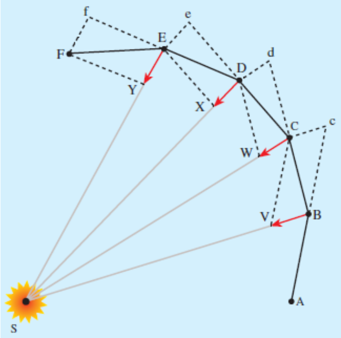

Figure 1 – According to Newton, of its own accord a body would move uniformly straight forward, and forces cause deviations from such states.

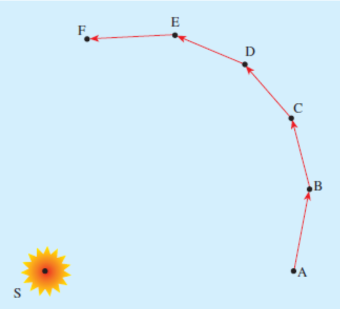

Figure 2 – According to Kepler, of its own accord a celestial body would remain at rest, and forces cause deviations from this state.

A similar explanatory strategy can be seen in the explanation by 'the man in the street' of why a cyclist has to keep pedalling in order to maintain speed. The simple answer is that if the cyclist stopped pedalling, he/she would come to a stop. In this scheme, the force takes the form of a personal influence (pedalling) and the force-free motion is the motion without this influence, i.e. to gradually come to a stop.

The explanations of motion by Newton, Kepler and the man in the street all have the same basic structure: some assumption regarding the force-free motion, in combination with some assumption regarding force laws to account for deviations from this force-free motion. Note in particular, as the cases of Kepler and Newton clearly illustrate, that under different assumptions about the force-free motion, different forces are required to account for the very same motion (see Klaassen, 2005, for further details).

The explanatory scheme outlined above can be seen as an example of how an educational designer can benefit from a thorough study of science, its philosophical foundations and its historical origins.

Design Principle: Tapping Core Causal Knowledge

The basic explanatory scheme, and the core causal knowledge in which it is grounded, can serve to direct and guide students' learning about how to explain the motions of bodies, i.e. it can be made to function as a kind of advance organiser (Ausubel, 1968). Appropriately tapping core causal knowledge can therefore be seen as a promising design principle, in order to provide students' learning about mechanics in advance with a sense of purpose and direction. Students command this core causal knowledge, though implicitly and most likely in a weakly articulated form. But the relevant intuitions can be stirred up in students by means of an appropriate exemplary case. It is here that Comet Kirch comes in (Figure 3). Our idea was to slightly modify the figure, and to present to students only the observed positions of the comet relative to the sun. Our expectations were, first of all, that this would present an intriguing case for students. It immediately raises all sorts of challenging questions. Where was the comet located at the dates it was not observed? What was its orbit? And above all, what exactly did happen between late November and early December 1680, and why?

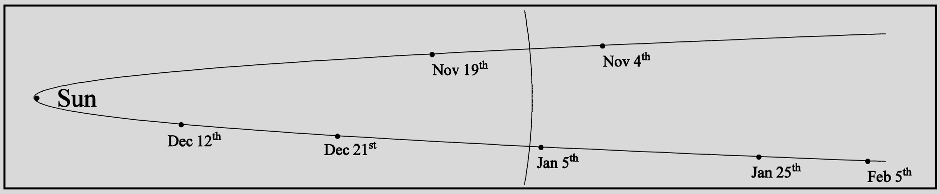

Figure 3 – Trajectory of the Comet Kirch, late 1680 and early 1681, adapted from Newton's *Principia* (Newton, 1999, p. 916). The circular arc denotes the radius of the earth's orbit.

Furthermore we expected that the case of Comet Kirch would stir up the relevant intuitions in the form a rich variety of student explanations, allowing the teacher to elicit the basic explanatory scheme as a subsequently useful guideline for further investigation. Let us somewhat elaborate.

In whichever way the comet will have moved precisely – some students may expect it to have turned in front of the sun, others may expect it to have turned around the sun – it surely must have made some kind of rather sharp U-turn. It is very unlikely that this is just how comets happen to move. However vague students' ideas may be about the way comets do happen to move of their own accord, surely not this way – this must be caused. It is also very easy to point at a plausible candidate responsible for causing the U-turn. The sun will have got something to do with it, whatever different ideas students may happen to have about how the sun did it. Furthermore, if we look at the comet's motion between December 21st 1680 and January 5th the next year, was the comet then still under the influence of the sun? Some students may suggest it must have been, because otherwise the comet would have come to a stop. Others may suggest that the comet was then so far away that the sun no longer could have had any grip on it; the comet was then still moving on because… well, comets do keep their speed even when left by themselves.

At first sight the task of the teacher to somehow manage this great variety of ideas into a useful guideline would seem nearly impossible, but this is where the explanatory scheme comes in. Underlying all the variety a basic element can be recognised: where the motion of an object differs from how we think it would have moved of its own accord, we go searching for an influence that may have caused the deviation. Even though at this point students are expected to see how to fit their intuitive reasoning into an explanatory scheme, an important problem still remains. How to determine which of the many varied combinations of assumptions about force-free motion and associated force laws is the right one? In answering this question the basic explanatory scheme is also expected to serve as a directive guideline. It provides students with a structure to work with, and they know what types of element they need to fit into the structure. In particular, the variety of assumptions that students are expected to have brought forward also prepares for the later introduction and comparison of the theories of Kepler and Newton for planetary motion, as alternative elaborations of the explanatory scheme.

The activity of The Comet Kirch can thus be seen as an example of a Problem-Posing Approach, because, when investigating the theories of Kepler and Newton, the students will know, on content-related grounds, what they are doing and why: 'fitting' both theories into the explanatory scheme, and finding out which of these theories serves best to explain the motion of comets, planets, moons and so on.

Discussion

The case of the Comet Kirch was expected (and in the end proved) to suit our educational aims and tenets. But the 'discovery' of this case was definitely not as straightforward as the above account may suggest. In previous versions of the teaching-learning activity we used the example of an asteroid on a possible collision course toward the earth. One of our reasons was that we expected students to appreciate this example as a case in which a call for prediction or explanation is important and challenging, given the severe consequences of an actual collision. In this respect we were right. We also believed, however, that the asteroid example would be useful to trigger useful intuitions about causal explanation to later further build on. In this respect we were wrong. When asked how to go about in predicting the asteroid's future course, the kind of explanation that was triggered was statistical instead of causal. Students did not suggest to go looking for other celestial bodies that might influence the asteroid's future course, but rather suggested to go looking for data about other asteroids. This suggestion is perfectly valid, and it does not show that students do not have the requisite causal intuitions. It is just that the asteroid case does not at that time serve to trigger those intuitions. Perhaps the students simply had the idea that the motion of the asteroid is not influenced by external objects at all: it just moves like asteroids happen to move, so let's see how other asteroids did move.

This left us with the problem of finding an appropriate example to trigger useful intuitions about causal explanation in the context of celestial motion. We searched the internet and browsed the tables of contents of astronomical journals, using search terms such as 'orbit determination'. We learned quite a lot, but nothing useful was coming forward. Months passed by, until we hit upon an article in which, as a side remark, mention was made of a method that Newton has used to determine the course of a comet – which led us to the Comet Kirch.

The search for an appropriate activity to stir up intuitions about causal explanation is an example of finding local solutions to local problems. Finding the picture of the trajectory of the Comet Kirch in the *Principia* was due to persistence and sheer luck.

Conclusion

During the whole process of searching in vain we were aware of the design principle of tapping core causal knowledge, but this awareness did not help to generate a concrete way of doing this. The function of the design principle was rather one of recognising, or checking, whether or not a particular activity fitted the basic tenet of our problem-posing approach. The design principle made us receptive to recognise a promising case for the initial teaching-learning activity. We do not wish to downplay this role. But our main point here is to illustrate that design principles are of little use when it comes to designing concrete teaching-learning activities. The example also serves to illustrate how much one is dependent on the goodwill of Lady Fortune. After all, one can be receptive and not receive. In the end we were lucky enough to receive.

4 Putting a Design to the Test

Adjusting a Design

When a design is put to the test, the theoretical predictions offered by the internal justification often do not come out. This does not prove that, after all, common-sense psychology was not the appropriate theoretical input. It is rather common-sense psychology that helps to lay bare the deviations from what was expected. The deviations are merely a reflection of the complex nature of a didactical design. A wrong estimation of students' initial attitudes, for example, may make the expectations go off-track from the very beginning and, as the errors accumulate, make them deviate increasingly in the further process.

In some cases it may be possible to 'explain away' a deviation from the predicted path. This may happen if the teacher did not guide the activity as intended, while there are indications that, had the activity been guided as intended, students would after all have done what they were expected to do. In such a case, one may decide not to make any adjustments (Kortland, 2001, pp. 120-126). More frequently the deviations reflect a clear need to make adjustments, though typically it will not be so clear which adjustments will suffice. Since a design is a highly interrelated complex, a failure that clearly emerges in one area may just be a symptom of a problem elsewhere. Another aspect of the interrelatedness is that necessary changes in one area are likely to require changes in several other areas. There are no rules – we hope it is clear – that tell where adjustments are to be made and which adjustments are to be made.

Some further, and deeper, complexity may arise if one decides not to make adjustments to the design in order to better realise the process that one wanted, but instead to make adjustments to what one wants the process to be like. That is, one may feel a need to adjust one's learning goals or even one's educational tenets – either with regret or with renewed hope.

Change of Learning Goals

The reasons for changing learning goals may be practical (e.g. not attainable in the given time), or principled (does not seem to be attainable at all), or both. Again, there are no rules that tell which goals to replace with which other goals, or where to go looking for new goals. The example given below is taken from an earlier version of the mechanics course referred to in the context of Example 2.

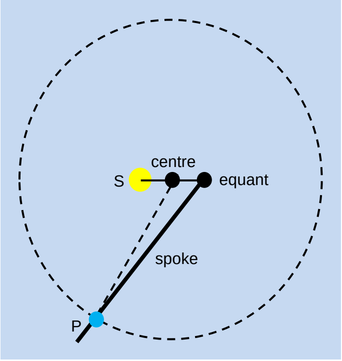

Figure 4 – Ptolomy's equant scheme. The planet P moves in a circle. The sun is a bit off-centre. The spoke has uniform angular motion with respect to the equant. The scheme yields an almost perfect geometrical-kinetic description of planetary motion, but not a *causal* explanation.

Example 3: Planetary Motion

A Historical Approach

In an earlier version of our mechanics course we tried a historical approach. We tried to make the need for causal explanation of planetary motion apparent by first elaborately discussing some earlier accounts of planetary motion. In particular, the Hellenistic account in terms of uniform motion in a circle. The geocentric accounts of Ptolemy in terms of cycles (for the sun and the moon), epicycles (for the planets), and epi-epicycles (for the moons of the planets) are well-known.

What is less known is that Ptolemy was well aware of some inaccuracies in his cycle-epicycle account. He knew that the distance between the sun and a planet is not exactly constant. Furthermore, the motion of a planet is not exactly uniform. Where it is closer to the sun, it moves faster. Ptolemy invented an ingenious scheme to account for these inaccuracies, by which he tried to uphold the Hellenistic ideal of circles and of uniform motions as much as he could. Put in heliocentric terms, the scheme is as follows (Ptolemy of course discussed it in geocentric terms). The planets still move in circles, though the sun is no longer exactly in the centre. But, and this is the ingenuity of Ptolemy's scheme, if we consider a point just as far from the centre as the sun but at the other side, then with respect to *this point* the angular motion of the planet is uniform. So if we imagine a spoke in this opposite point, and turn the spoke uniformly around, then this is how the planet moves along the circle, whilst sliding up and down the spoke (see Figure 4). Ptolemy called this special point opposite to the sun the *equant*, or equalizing point. To recapitulate, there is still a circle, but the sun is no longer in the centre; there is still uniform angular motion, but with respect to the equalizing point. The predictions of this model are in near perfect agreement with observations. (The reason why the scheme works so well is that the eccentricities of the planetary orbits are relatively small. We refer to Barbour (2001, section 3.7) for further details.)

In his 1609 book *Astronomia Nova* (*New Astronomy*), Kepler applauded Ptolemy for the ingenuity of the equant scheme – and indeed it is a marvellous achievement, bearing in mind that it is entirely based on naked eye observations. But despite all his admiration, Kepler at the same time also began to pose some serious questions. Why – and we can almost hear Kepler scream – why are you paying so much attention to that equalizing point? It's a void; there's nothing there; there is no spoke; this is all imaginary! But look here, just around the corner, relatively speaking, there is this real thing – a gigantic star! Is it not far more plausible that the sun has got something to do with the planet's motion, rather than all this imaginary stuff? What is more, so Kepler went on, is that this would also explain why a planet moves faster the nearer it is to the sun. For the nearer to a cause, the greater the effects are. Furthermore, all the planets revolve in the same sense around the sun, and this sense is also the sense in which the sun rotates about its own axis. The same is the case for the earth and the moon. The moon revolves around the earth in the same sense as the earth rotates about its own axis. This cannot be accidental. Everything added up for Kepler. Somehow the sun, or its rotation, must be the cause of planetary motion; just as the earth, or its rotation, must be the cause of the motion of the moon. Here we have the birth of the physical concept of force. Not Newton is the father; Kepler is (see Stephenson, 1994). Or perhaps it is better to opt for co-parenthood. Newton's essential innovation was to take *uniform rectilinear motion* as the state of motion, where Kepler had taken *rest* (see Figures 1 and 2).

A Change of Learning Goal

What we intended to induce in students, by a carefully chosen sequence of tasks, was a similar sort of enthusiasm for *causal* explanation of planetary motion as Kepler so clearly felt. In the test it turned out, on the one hand, that the approach was promising, but, on the other hand, that the activities took considerably more time than we had anticipated. In fact, keeping them in would make the course as a whole cost much more time than the teachers involved had available. The Hellenistic way of accounting for planetary motion is no longer part of the current version of our mechanics course, though we still think it's a good idea that deserves to be elaborated.

Change of Educational Tenets

The test of a design may also give rise to a change of educational tenets. An example is a study in which it was tried to make students experience, in so-called conceptual practicals, that their alternative conceptions are untenable (Dekkers, 1997). In the end Dekkers abandoned this educational tenet, because it turned out that most of what students said and concluded while doing the experiments was simply correct. To be more precise, Dekkers found that if he no longer assumed that students mean by physical terms what a physicist means by those terms, that then he could give a consistent interpretation of students' words such that what they said was largely correct. By obeying a sound methodological principle – to so interpret others that what they say comes out largely correct – the original educational tenet of Dekkers evaporated.

The Object of Empirical Research

The only direct object of empirical research is the internal justification of a didactical design. Testing an internal justification is independent of the value-laden choices that are adopted, and can be done without knowing what these choices are. It is 'just' a matter of establishing the extent to which the teaching-learning process proceeds as hypothesized. Nevertheless the findings may indirectly lead to adjustments in value-laden choices as well, e.g. when one unexpectedly uncovers a new goal as potentially worthwhile – i.e. ontological innovation in the sense of diSessa & Cobb (2004). Or occasionally one may come to feel that one has been led astray by one or another orienting framework. But such side-effects, even when welcome, cannot be the point of design research.

Lack of space prevents us to elaborate in detail an example of what we consider to be a case of ontological innovation. Let us just briefly mention that, based on an analysis of students' reactions to the series of tasks concerning the Chernobyl Accident (see Example 1), we began to seriously doubt our original assumptions, expectations and aims concerning these tasks. In particular we have come to realize that it would have been a much more appropriate aim to just make pupils come to clearly recognize, and appreciate as a real problem that is worth solving, something like this. Under which conditions can where radiation be measured? In this sense the results of the try-out of the first version of our teaching-learning sequence on radioactivity have lead us to ‘discover’ the basic tenet of our problem-posing approach. For further details we refer to Klaassen (1995, section 7.4).

5 The Nature of a Tested Design as an Empirical Theory

A ‘Good Enough’ Design

Empirical support for a didactical design is a claim of the following kind: If handled with proper care, the teaching-learning process will proceed more or less as intended, under normal circumstances. The aim of improving a didactical design cannot be to eventually arrive at ‘the ultimate’ design – one whose predictions will come out in exactly the predicted way. All that matters is that a design can be judged ‘good enough’ to serve as a valuable guideline for understanding and guiding what goes on in actual classrooms. In each actual case the teaching-learning process will without doubt meander in a somewhat different way around the main predicted path. Several revisions are typically needed before one is even willing to consider the question whether or not a design can be judged good enough, and the first revisions most likely require considerable adjustments. But no matter in how many classes or with how many teachers one has tested a design, the claim that it is good enough will always be hedged with unspecified clauses such as ‘proper’, ‘more or less’ and ‘normal’. No further specification is needed, however, because generality is *not* the aim of testing designs. A design is expected to work in new circumstances, not because it worked in similar circumstances, but because this is how rational creatures are expected to behave.

Educational Relevance of Tested Designs

In our opinion, internal justification ought to play a pivotal role in design research. In many design experiments we find an internal justification and its detailed test wanting. This severely undermines the educational relevance of such experiments.

The theoretical yield of design research consists in nothing but domain-specific theories in the form of tested designs. There is every reason to rest content with this. Tested didactical designs provide just the right kind of evidence-based contributions to the expertise of both teachers and researchers, and leave ample scope for progress and comparison.

The explicit specification of the value-laden choices and the detailed account of the envisioned teaching-learning process allow a teacher to get a feel of whether or not the process appeals to him. In combination with the empirical support the teacher can form a judgment as to whether or not he can make it work in his circumstances, or see himself able to adapt it to his specific circumstances. In this sense a good enough design allows a teacher to reach an informed decision about whether or not to make an effort to use it.

Progress in science education research may occur in at least three ways. First, within the quality standard set by given learning goals and educational tenets, one design may arguably *better* meet the standard than another. Second, within the quality standard set by fixed educational tenets, for a *growing* number and variety of topics (with associated learning goals) one may be able to produce good enough designs. Third, researchers operating with different quality standards can *critically discuss* the ways in which their respective theoretical perspectives have differently shaped the concrete activities in their respective teaching-learning sequences. An example of such a critical discussion, in this case concerning a teaching-learning activity called ‘The BIG Circuit’, is given below. The activity is part of an introductory electricity course for students in lower secondary education.

Example 4: The BIG Circuit

Design Principle: Learning Demand

The design principle ‘Learning Demand’ (Leach & Scott, 2002) directs a designer within the general framework of a ‘Sociocultural Approach’ to follow this procedure: (1) analyse the school science knowledge to be taught in terms of ontology, epistemology, and the patterns of reasoning on which explanations are based; (2) analyse how this area of science is conceptualised in everyday social language (in the same terms); (3) “identify the learning demand by appraising the nature of any differences between 1 and 2” (Leach & Scott, 2002). Learning demand draws upon the theory of social constructivism in that it involves making a comparison between two social languages: the social language of (school) science and the everyday social language that students are likely to use. What gives Learning Demand its point, as a design principle within a social constructivist approach, are the basic tenets associated with social constructivism: there are essential differences between the two social languages with respect to concepts, associated ontology, epistemology and/or patterns of reasoning; consequently “learners must reconstruct the sense of the talk and activities that surround them on the social plane, reorganizing their existing ideas and ways of thinking accordingly" (Ruthven et al, 2009). Given sociocultural theory, Learning Demand is an obvious design principle.

An example of a teaching-learning activity that is informed by Learning Demand is presented in Ruthven et al (2009); the quotes in the following discussion refer to this article. The activity is part of a teaching-learning sequence designed to address a learning demand about the patterns of reasoning used to explain the behaviour of simple electrical circuits. In particular, the social languages of students and school physics are claimed to be based on different patterns of reasoning, as follows: "students' explanations tend to be based on a linear causal sequence of events, starting in the battery with events in the resistive components of the circuit following later. By contrast, the social language of school physics describes circuits as integrated systems where events happen at the same time." The activity to address this learning demand is then described as follows.

> "The teaching activity begins with an activity called the BIG Circuit. Prior to the lesson, the teacher constructs a simple series circuit with a power source at one end of the teaching room, a bulb at the other end of the teaching room, and very long wires stretched around the walls of the room connecting the circuit together. Pupils are asked to predict what will happen when the circuit is connected. Many will say that *there will be a short delay* between connecting the circuit and the bulb lighting. This is because they believe that *the electricity will take a short time* to travel from the power source to the bulb. However, when the circuit is connected there is no delay: The lamp lights instantaneously. This raises a question to be addressed in subsequent teaching: *How come the bulb in the BIG Circuit lights instantaneously*? […]
> The next stage of the activity sees the teacher introducing an analogy to the class. The particular analogy involves bread vans (i.e., charges) delivering bread (i.e., energy) from a bakery (i.e., power source) to a supermarket (i.e., bulb). The bread vans are in a nose-to-tail loop all the way from the bakery to the supermarket so that when one bread van stops, all have to stop, and when one moves, all move. This explains why as soon as the bread vans move (i.e., the circuit is connected), bread can be delivered (i.e., the bulb can light)." (p.340).

In short, the activity is meant to make students reject that there is a delay between events in the power source and events in the resistive components, and to withdraw their application to this case of the pattern of reasoning that consists of a linear causal sequencing of events. Instead they are to accept the school physics pattern of reasoning in which the description of circuits is such that all events (in the analogy: the motion of the bread vans from bakery to supermarket, the delivery of bread) are happening at the same time.

Discussion

Let us first of all formulate some points where we agree. Students' expectation of the delay is indeed based on linear causal sequencing, and the first part of the activity may make students believe that there is no delay.

But there are also points where we disagree. According to us, first of all there is nothing wrong with applying linear causal sequencing to the BIG Circuit. And secondly, the analogy does not address the issue whether or not there is a delay, and if further elaborated reinforces the idea that there must be a delay.

Why do we disagree? One of the main lessons to be learned from the many studies on students' ideas about electricity is that lower secondary students' pre-knowledge about circuits is practically zero. Whatever sense students can make of circuits will be guided by general patterns of causal reasoning. We believe that their expectations concerning the BIG Circuit will be based on roughly the following reasoning.

> If the circuit was connected without the battery, the bulb would not have lit. Probably the battery causes the bulb to light. Since in the BIG Circuit the bulb is quite some distance away from the battery, there will be a time delay between the events starting in the battery when the circuit is connected and their ultimate effect (the bulb lighting). Probably the delay will have something to do with mediating events that take place in what connects the bulb to the battery (the wires).

Note that this account involves the identification of a causal relation, the idea that there will be a time delay between cause and effect if the effect takes place some distance away from where the cause happened, and the idea that in such a case cause and effect will be linked up by a chain of spatially and temporally contiguous events. As already noted, the account is of a very general nature. Students will have little or no idea about the nature of the processes that originate in the battery when the circuit is connected, about how the battery needs to be connected to the bulb (will one wire do?), about the nature of the intermediate processes in the wires, and so on.

But granted that there are many gaps in students' account when it comes to the details, which of course are to be expected for lower secondary students, we do not see that there is much wrong with the general account itself. Though scientists are able to provide considerably more detail than students, they will proceed along the same general lines when thinking about what will happen after the circuit is connected. Space does not allow us to give a full account of the physics involved, but at least part of the account boils down to a rearrangement of surface charge on the wires and a resulting electric field growing rapidly outward at approximately the speed of light from the region where the circuit is connected (see, for example, Chabay & Sherwood, 2011, section 19.5). This reasoning of scientists does not differ in kind from the reasoning of students, although it does of course employ a lot of scientific theory that will be completely alien to students. Moreover, if only asked whether or not in the BIG Circuit there will be a delay between connecting the circuit and the bulb lighting, scientists will all agree that the bulb does *not* light instantaneously – that in fact there will be a time delay, though so short (about 10^{-8} s) that it cannot readily be noticed in the BIG Circuit. However, the delay would be noticeable (about 1 s) in a REALLY BIG Circuit – say, when connecting a power source on Earth with a bulb on the Moon.

When the BIG Circuit activity evolves as planned, students are made to conclude that there is no delay – that the bulb lights instantaneously – and will get the impression that there is something wrong with the reasoning that led them to expect a delay, thus making them susceptible to the analogy to be introduced subsequently: “The bread vans are in a nose-to-tail loop all the way from the bakery to the supermarket so that when one bread van stops, all have to stop, and when one moves, all move. This explains why as soon as the bread vans move (i.e., the circuit is connected), bread can be delivered (i.e., the bulb can light).”

Our criticism is twofold. First, as argued above, there is nothing wrong with the general reasoning that leads students to expect a delay, and consequently there is also no need to make students experience that there is something wrong with their reasoning. Secondly, ignoring the first criticism and going along with the idea that students’ reasoning needs to be supplanted by a superior kind of reasoning, the analogy offered does not in fact provide a superior reasoning. The analogy does not even address the issue whether or not there is a delay. Instead it simply takes for granted that when all the vans are placed nose-to-tail (i.e., when the circuit is connected), that then the vans (i.e., the charges) move right away. However, it then is still hard to understand that bread will be delivered right away to the supermarket (i.e., that the bulb will light instantaneously). If we go along with the analogy, the vans initially near the supermarket will still be empty and have no bread yet to deliver. It is only when the vans that were initially near the bakery (and there were uploaded with bread) eventually reach the supermarket – and this will take some time – that the downloading of bread to the supermarket can begin. The only way out of this prediction of a delay would be to supplement the analogy to the effect that somehow all the vans between the bakery and the supermarket already contained bread. But this would deprive the analogy of its initial intelligibility. If anything, the analogy reinforces the prediction of a delay.

Conclusion

The inappropriate use of the BIG Circuit activity, in our view, can be explained by the designers’ educational tenet of Learning Demand as a design principle, in line with their sociocultural approach. This strongly directed them to go looking for essential differences between the targeted scientific way of reasoning and some pattern of reasoning associated with students’ colloquial language. In order to satisfy this theoretical demand, the designers identified linear causal sequencing as the everyday candidate. They then had to think of an activity that would make linear causal sequencing somewhat abhorrent to students. This is where they thought the BIG Circuit would be of service, because linear causal sequencing supports the prediction that there is a delay in the BIG Circuit, whilst the students must be made to conclude that there is *no* delay. In our opinion we here have an example where an educational tenet (Learning Demand) and its underlying theory (Sociocultural Approach) orient the designer in a completely wrong direction (as stated in Section 3).

Despite our criticism of the BIG Circuit activity, we applaud the designers for making this criticism possible. They presented, at a fair level of detail, what educational work they expected the BIG Circuit activity to do. This allowed us to indicate, first, that there is no need for such educational work and, second, that the activity simply cannot deliver this work.

Exchange of ‘Good Enough’ Designs

Discussions such as the one about the BIG Circuit activity, in which researchers critically and in detail examine the workings of each other’s educational designs, ought to have a much more prominent place in the science education research literature. At least this may lead to clarification of the frameworks at stake. Above all such an exchange will keep theoretical considerations firmly secured to what they are supposed to be relevant for: concrete teaching-learning activities. This is progress too, when compared to the abstract and freewheeling manner in which theoretical frameworks are frequently discussed in the literature.

Without doubt the researchers in direct critical contact will not be able to convince one another, but for other researchers the exchange may clearly bring to the fore the approach of either contestant. By providing telling examples, supported by cogent reasoning involving design principles, the exchange may in a clear and succinct way draw the attention of researchers to an easily overlooked educational aspect, convince some researchers of the importance or value of that educational aspect, sensitize other researchers to basic tenets of one approach or another, and perhaps even incite some researchers to adopt one approach or another. In all of this, design principles can play a useful clarifying role. But what in the end counts is whether or not a concrete teaching-learning activity does the educational work it is supposed to do – whether or not the internal justification of the activity is sound. This soundness can be judged without knowing the design principles that informed the activity. Conversely, whatever comes out of the test of a design has no direct bearing on design principles.

6 Conclusion

We have argued that, in our view, educational theory consists of internally justified and empirically supported designs. Useful as design principles may be for the clarification of one's approach, they are not part of educational theory nor do they facilitate educational researchers to devise new educational designs. It is rather by being exposed to a lot of educational designs and experimental tests of such designs that science education researchers – or at least the talented ones – will be able to contribute empirically testable theories or experimental tests of such theories. As a scientific community we should produce and test as much designs as we can, and critically discuss each other's designs in order to make progress. We are under no compulsion to deliver anything over and above specific educational designs and experiments, though on occasion one may do so if one thinks it will help to clarify what one has done. But it is wrong to think of design principles as the truly worthwhile outcome of science education research, by virtue of which the design process will be made easier or more efficient.

The production of good enough designs does not sit easy with the current emphasis on quantity of 'output'. It takes quite a lot of effort and time to produce such a design, and because of its considerable complexity it is hard to report concisely its justification and its test. This puts serious pressure on the progress in what we see as the core business of science education research: to construct, critically discuss, and empirically evaluate detailed content-specific justifications of teaching-learning sequences.

Acknowledgement

We wish to thank Piet Lijnse. His writings on developmental research and didactical structures (see Kortland & Klaassen, 2010) have guided and inspired us to develop the view on the role and nature of theory in design research outlined in this paper. We are quite sure, though, that he would sensibly criticize our view in his customary friendly way.

References

Ausubel, D.P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.

Barab, S., & Squire, K. (Eds.) (2004). Special issue on design-based research. *The Journal of the Learning Sciences, 13*(1).

Barbour, J. B. (2001). *The Discovery of Dynamics: A study from a Machian point of view of the discovery and the structure of dynamical theories*. Oxford: OUP.

Chabay, R.W., & Sherwood, B.A. (2011). *Matter & interactions*. Hoboken, NJ: Wiley.

Davidson, D. (2004). *Problems of rationality*. Oxford: Clarendon Press.

Davidson, D. (2006). The perils and pleasures of interpretation. In E. Lepore & B.C. Smith (Eds.), *The Oxford handbook of philosophy of language* (pp. 1056-1068). Oxford: Clarendon Press.

Dekkers, P.J.J.M. (1997). *Making productive use of student conceptions in physics education – Developing the concept of force through practical work*. Amsterdam: Free University.

De Klerk, L.F.W. (1982). Verslag van het AERA-Congress 1982 te New York. *Pedagogische Studiën*, *59*, 519-21. Report, in Dutch, of the AERA-Congress in New York in 1982.

DiSessa, A.A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *The Journal of the Learning Sciences, 13*, 77-103.

Emmett, K., Klaassen, K., & Eijkelhof, H. (2009). Introducing mechanics: Towards a justified choice between alternative theories of motion. *Physics Education, 44*, 492-498.

Klaassen, C.W.J.M (1995). *A problem-posing approach to teaching the topic of radioactivity.* Utrecht: CDβ Press. To be retrieved from http://igitur-archive.library.uu.nl/dissertations/01873016/inhoud.htm

Klaassen, K. (2005). The concept of force as a constitutive element of understanding the world. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof (Eds.), *Research and the quality of science education* (pp. 447–457). Dordrecht: Springer.

Klaassen, K., Westra, A., Emmett, K., Eijkelhof, H., & Lijnse, P. (2008). Introducing mechanics by tapping core causal knowledge. *Physics Education, 43*, 433-439.

Kortland, J. (2001). *A problem-posing approach to teaching decision making about the waste issue*. Utrecht: CDβ Press.

Kortland, J., & Klaassen, C.W.J.M. (Eds.) (2010). Designing theory-based teaching-learning sequences for science education. Utrecht: CDβ Press. To be retrieved from http://www.staff.science.uu.nl/~kortl101/book_sympPL.pdf

Leach, J., & Scott, P. (2002). Designing and evaluating science teaching sequences: An approach drawing upon the concept of learning demand and a social constructivist perspective on learning. *Studies in Science Education, 38*, 115-142.

Méheut, M., & Psillos, D. (Eds.) (2004). Special issue on teaching-learning sequences in connection to the aims and tools for science education research. *International Journal of Science Education, 26*(5).

Millar, R. (1989). Constructive criticisms. *International Journal of Science Education, 11*, 587-596.

Newton, I. (1999). *The Principia: Mathematical principles of natural philosophy*. Berkeley: University of California Press. New English translation, based on the third (1726) edition of *Philosophiae Naturalis Principia Mathematica*; the first edition was published in 1687.

Ruthven, K., Laborde, C., Leach, J., & Tiberghien, A. (2009). Design tools in didactical research: Instrumenting the epistemological and cognitive aspects of the design of teaching sequences. *Educational Researcher, 38*(5), 329-342.

Stephenson, B. (1994). *Kepler's Physical Astronomy*. Princeton: Princeton University Press.

Tiberghien, A. (2000). Designing teaching situations in the secondary school. In R. Millar, J. Leach, & J. Osborne (Eds.), *Improving science education: The contribution of research* (pp. 27-47). Buckingham: Open University Press.

Viennot, L. (2003). Relating research in didactics and actual teaching practice: Impact and virtues of critical details. In D. Psillos, P. Kariotoglou, V. Tselfes, E. Hatzikraniotis, G. Fassoulopoulos & M. Kallery (Eds.), *Science education research in the knowledge-based society* (pp. 383-394). Dordrecht: Kluwer Academic.