A Forceful Adventure: the use of interactive fiction to promote conceptual change.

Abstract

In recent years, researchers within science education have started to consider the impact of narrative upon teaching and learning in science. This paper investigates the possibilities of interactive fiction as a means by which pupils can be provided with feedback on their understanding in science, and explores the mechanisms which might allow learning from this. Through a review of literature around the use of narrative in science education, we produce a list of recommendations that might guide the development of interactive fiction within science education. These recommendations are tested through a small-scale study in which an interactive fiction book was written around Newton's laws, and then tested with 27, 16 and 17-year-old chemistry students, eight of whom also study post-compulsory physics. The interactive fiction developed is based upon the well-established Force Concept Inventory, and this allows progression of student understanding to be analyzed. The study finds that, upon reading the book, there is a significant positive effect size on the understanding of students who do not study advanced physics. The gains for those who do study post-compulsory physics were not statistically significant. The participants' report of enjoyment also discussed.

1. Sense-making through narrative

This paper begins with a review of research on the role of narrative in learning, and in learning science in particular, during which eight recommendations are made for the incorporation of narrative into interactive fiction during the teaching of science. The paper then considers a small scale study in which these recommendations were used to develop and test an interactive fiction book which diagnoses and addresses pupils' naïve concepts around forces.

In *Actual Minds, Possible Worlds*, Bruner (1986) argues that there are two modes of knowing employed by people when trying to make sense of the world they live in. The first, which he terms 'paradigmatic', concerns logical thinking and involves conceptualisation such that a number of these concepts can be related together 'to form a system' (Bruner, 1997, p12). This system is the 'instrument of reason' (Bruner, 1987, p2) and tallies with a scientific approach to the world. The second he pronounces the 'narrative mode', which refers to how 'we organize our experience and our memory of human happenings mainly in the form of narrative – stories, excuses, myths, reasons for doing and not doing, and so on' (Bruner 1991, p4). Thus, Bruner suggests that:

> It is very likely the case that the most natural and the earliest way in which we organize our experience and our knowledge is in terms of the narrative form.
>
> (Bruner 1996, p 121)

Following Bruner's lead, researchers within science education have started to consider the impact the above may have on possible new methods of teaching science, particularly with respect to using narrative (e.g. Millar & Osborne 1998; Norris et al. 2005; Avraamidou & Osborne 2009; Allchin 2014; Klassen & Klassen 2014). One obvious way of employing narrative within the science classroom is through books, and we will show that interactive fiction books have the potential to meet several of the recommendations made in the literature in relation to the educational use of narrative. As such, Section 1 of this paper considers the influence of narrative on learning in science: Section 1.1 considers what research has to say about narrative, and Section 1.2 considers how this integrates with the teaching of a specific content area: Newton's laws of physics. Section 1 will develop eight recommendations for the way that narrative might be used within science education, with a particular focus upon interactive fiction. Section 2 will then report on a small scale study in which the impact of reading a narrative fiction book upon conceptual understanding is tested.

1.1 Narrative in science education

In their report *Beyond 2000: Science Education for the Future*, Millar and Osborne (1998) issued a clarion call to employ the narrative form more often when communicating scientific ideas, in order to make them much more 'coherent, memorable and meaningful' (Millar & Osborne, 1998, p2013). This call went on to have considerable influence on curriculum developments in the UK and promoted seeing science as the development of 'explanatory stories' through which pupils can increase their understanding of the world. Yet some ten years later, Avraamidou & Osborne (2009) echoed Lemke's concern around a perceived and pernicious 'mystique of science', which makes 'science seem dogmatic, authoritarian, impersonal, and even inhuman' (Lemke, 1990, p xi). Avraamidou & Osborne extended this, postulating that students stand outside looking in on the theories they learn, the net result being disengagement and even alienation with the subject. They contended that there are four principle types of text that are typically employed to communicate science. These are: argumentative; expository; narrative; and finally, those that combine narrative and exposition. Argumentative texts are dialectical in nature, starting from a particular premise and leading to a conclusion that is often hedged and debatable. A study by Penney et al. (2003) found that science textbooks do not use this type of text at all, despite the importance of argumentation within science education. Exposition is the form that traditional textbooks take, still a staple diet regarding the communication of science within many classrooms. These are ‘univocal [and] non-dialectic’ (Avraamidou & Osborne, 2009, p1685) and are intended simply to describe or explain.

This expository form is at odds with the fact that ‘everyday discourse is narrative’ (Avraamidou & Osborne, 2009, p1686). In trying to answer what ‘narrative’ is, Avraamidou and Osborne (2009, 1688) focus on three analyses of narrative in particular (Chatman, 1978; Toolan 2001; Norris et al. 2005) and conclude their meta-analyses with a table of key constituents of narrative coupled with their function, reproduced as Table 1 here.

Table 1 – key components of narrative

Narrative component	**Description of component**
Purpose	To help us understand the natural and human world. In the case of the natural world, narratives help the reader to invent new entities, concepts and some picture of the scientist’s vision of the material world
Events	A chain or sequence of events that are connected to each other
Structure	An identifiable structure (beginning, middle, end) where events are related temporally
Time	Narratives concern the past

Agency	Actors or entities cause and experience events. Actors may either be human or material entities who act on each other
Narrator	The teller who is either a real character or alternatively, a sense of a narrator
Reader	The reader must interpret or recognize the text as a narrative

(from Avraamidou and Osborne 2009, p1693)

The inclusion of these narrative components is important in developing forms of narrative that overcome the separation of pupils from the theories and ideas they learn in science. As such, this forms the first recommendation for the use of narrative in science education.

Recommendation 1: Narrative should incorporate Avraamidou and Osborne's (2009) necessary components: purpose; events; structure; time; agency; narrator; reader.

Narrative effects and the privileged status of story

Willingham (2004; 2009) has described stories as being 'psychologically privileged'. He cites Britton et al. (1983) who relate 'cognitive capacity' in reading to the type of text provided. They found that 'capacity was filled more completely while reading narrative text than while reading expository text' (Britton et al., 1983, p39) and that people find stories intrinsically more interesting than expository text, the form of text Avraamidou & Osborne (2009) declared to be the most commonly used in the science classroom. Comprehension is also affected by type of narrative. ~~Bets~~ Best et al.'s (2008) study of third grade elementary school children showed a greater comprehension of narrative than expository text, although prior knowledge influenced this. Duke (2000) found that a difference in how much students at this age have been exposed to the two different types of text determined the relative influence of different text types. This followed an earlier study by Englert & Hiebert (1984) which demonstrated that ‘knowledge of discourse types underlies effective expository comprehension’, with this ability increasing as students get older. Furthermore, Graesser et al. (1994) discovered that narrative texts were read significantly faster than expository texts, which the researchers took to mean that they were easier to comprehend. This is supported by Zabrucky & Moore (1999) who found that, in addition to reading more slowly, adults ‘reread nonproblematic and problematic information in expository passages more frequently’ (Zabrucky & Moore, 1999, p706) and yet their ability to recall problematic information was poorer.

In addition, narrative texts have been shown to be easier to remember than expository texts (Graesser et al. (1994)) and that this is likely due to the causal connections individuals make when reading narratives (Keenan et al. 1987). These studies built upon earlier exploration of the psychology of narrative, notably Bartlett’s groundbreaking study on memory (1932/1995). He asked twenty Edwardian schoolboys to read the Native American Folktale ‘War of the Ghosts’. They were then asked to recall the story a number of times with the interval between retellings ranging from fifteen minutes to several years. Bartlett noticed that the retellings resulted in the story being ‘robbed of all its surprising, jerky and inconsequential form, and reduced … to an orderly narration’ (Bartlett, 1932/1995, p153). Bartlett thought of ‘every human cognitive reaction – perceiving, imaging, remembering, thinking, and reasoning – as an effort after meaning’ (Bartlett, 1932/1995 p44). Evidence that recall is improved by such ‘effort after meaning’ has subsequently been found (e.g. Auble & Franks 1978; Zaromb & Roediger 2009). Narrative has also been shown to influence the persuasiveness of information regarding a particular understanding (Voss, Wiley, & Sandak 1999), which is pertinent to science education (Norris et al. 2005). However, we should caution that there is likely a difference between a recall of narrative details and the application of scientific understanding or skills.

Applying the above findings to science education is not without potential problems however. In a review of the literature, Arya and Maul (2012), found that a mixing of genres (e.g. narrative and expository) could cause confusion in readers (Hidi et al. 1982), that the narrative had the potential to distract readers from the key content (Wade & Adams 1990) and, more worryingly, could lead to readers developing misconceptions (e.g. Jetton 1994; Madrazo 1997; Mayer 1995; Rice 2002). Perhaps unsurprisingly, this suggests that narratives are not, in-themselves, a force for good in science education but must be constructed and used in the right way. We will now consider what research has to say about reducing these potential issues in the use of narrative within science education.

Speaking to the reader

Arya and Maul (2012) suggest that two methods in particular are likely to benefit how students interact with texts (of any type) in science education: making the author visible and employing the personalization principle. Paxton (1997, 2002) discovered that adolescent students preferred texts which have a *visible author*, who writes in the 1st person, and reveals personal information and opinions. Students engaging with visible authors thought more about the topic, wrote longer essays, and used more rhetorical devices associated with personalization (e.g. use of personal pronouns such as ‘the way I see it…’). Paxton argues that this is because students ‘hold mental conversations with and about those authors’ (Paxton, 1997, p 235). These findings can be related to the ‘personalization principle’ which has been developed primarily within the field of multimedia and e-Learning (Clark & Mayer 2011) although much of its early research had taken place with respect to teaching science. Moreno and Mayer (2000, 2004) investigated the effects of employing a personalized style (use of first and second person perspectives) in contrast to a neutral style (use of third person perspective) when teaching middle-school science to psychology college students. Across seven studies they found that students exposed to the personalized style did better on follow-up transfer tests (with effect sizes all above 1). A further study compared the learning of students who watch a personalized version of a narrated animation that explained how the human respiratory system works with those who watched a formal style version. The difference in styles was simply the replacement of ‘the’ with ‘your’ on twelve occasions e.g. ‘During exhalation, the[/your] diaphragm moves up, creating less room for the[/your] lungs…’ (Clark & Mayer, 2011, p186-187). These simple changes resulted in an effect size of 0.79 regarding a transfer test. In line with Paxton’s findings, Clark and Mayer (2011) argue that a personalized voice gives students the impression of being in conversation with the author, which results in greater engagement and an increase in effort to understand. They suggest that a hybrid of the visible author and personalization principle can be achieved through the use of a pedagogical agent, which can be thought of as being a type of personal coach.

An example of this within science education is the ‘Design-a-Plant’ e-learning tool, which focuses on botanical anatomy and physiology (Lester et al. 1999). Herman, an animated pedagogical agent in the form of a bug, poses questions, offers advice and provides feedback to students as they design a plant to survive in specified environmental conditions. Moreno et al. (2001) ran a series of experiments involving college and 7th grade students (aged 18 and 12-13 respectively) in which students were randomly assigned to one of two groups: one assisted verbally by Herman and the other group receiving identical graphics and explanations without a pedagogical agent. Students in the former group reported greater interest in the activity and performed better on subsequent transfer tests. There was no difference in retention. Moreno et al. (2001) also found that having the text spoken by Herman proved more effective than simply presenting the words on-screen, although it is not clear whether this is to do with personalization or oral narration.

The strong evidence base for personalization and the use of pedagogical agents, leads us to make this a further recommendation for the development and use of narrative within science education:

Recommendation 2: Narrative should employ the personalisation principle and pedagogical agents whenever possible.

Interactive books

Much of the work around the personalisation principle and pedagogical agents is done in relation to e-learning and software. However, we were not able to find any research around the use of interactive fiction books in physical form, let alone their use within science education. Many of the principles which apply to interactive fiction using information technology undoubtedly apply to physical books, for example the personalization principle just discussed is an intrinsic feature within interactive books, and they also allow for the use of pedagogical agents through characters presented. Physical books have different characteristics to online resources however, for example in their physical manipulability (e.g. in moving backwards and forwards in a narrative) and in how pupils relate to books over online learning environments. Mangen and colleagues (Mangen et al 2013; Mangen & Kuiken 2014) found a difference in readers' comprehension of texts, and their narrative engagement with them, when the format of presentation is paper rather than on screen. The researchers did reading pretests in reading comprehension, word reading and vocabulary in a Norwegian school. A multiple regression analysis was carried out to investigate to what extent reading modality would influence the students' scores on the reading comprehension measure. Students who read texts in print scored significantly better on the reading comprehension test than students who read the texts digitally.

With this in mind we considered it worthwhile to investigate the impact of an interactive fiction book on conceptual understanding in science. Moreover, this study was prompted by pedagogical and practical reasons: interactive books have been published for more than 50 years and may be a familiar and enjoyable format to some pupils. They can also be easily deployed in homes and classrooms without the need for information technology. Furthermore, the writing and testing of a book was practicable for a classroom teacher (the first named author of this paper). Whilst the two recommendations made so far apply more broadly to narrative in science education, the discussion below thus makes recommendations in the context of interactive books specifically (and later their use in relation to learning about forces).

It is worthwhile considering firstly the history of interactive books, in order to draw on the experience of publishers to date. In the early sixties, a series of interactive 'scrambled' textbooks was published by Doubleday under the series title *TutorText* (Lipson 1962). Titles ranged from *Parliamentary Procedure* to *The Game of Chess* and included a number of science titles such as *Chemistry* and an *Introduction to Genetics* (Katz 2016). Readers were presented with information that they were then tested on via a multiple choice question, with the choice leading to specific page number. A correct answer takes them to the next piece of information, whereas an incorrect answer sends them to a page that elaborates further on the assessed topic and then asks them to return to the original question. The guiding principle is that the reader can only progress once they have successfully answered the previous piece (Crowder 1963).

Over the next couple of decades, 'programed textbooks' (Lipson 1962, p7) began to fall out of favour and would eventually be replaced by hypertext and multimedia. However, another type of interactive book was yet to reach its zenith, one using a fictional narrative. The 1980s saw an explosion of fantasy adventure gamebooks with the *Fighting Fantasy* series the most successful in the UK. Aimed primarily at children aged 9-12, the books sold in their millions (Green 2014). Written in the second person present, each *Fighting Fantasy* book begins with a background designed to set the scene of the adventure to comes. This is followed by a sequence of numbered sections, most of which end with a series of choices that lead onto another section. A reader's journey through the book is therefore determined by the option they select. The 1980s also saw a proliferation of interactive text-based computer games that involved graphical content. These include adaptations of novels such as *Treasure Island*. Lancy & Hayes (1988) research into the benefits of interactive fiction on 'reluctant readers' found that children labelled as such were often the most enthusiastic regarding interactive fiction.

This brief consideration of the history of interactive fiction firstly supports the suggestion that fiction is favoured (at least in commercial terms) over interactive textbook formats. It also suggests that the second person present tense: the text narrating the story and choices, is the accepted format of interactive texts. This leads us to our third recommendation:

Recommendation 3: Interactive narrative texts should be written in the second person present, in line with the commercially accepted format.

Explanatory versus corrective feedback

Since the *TutorText* series, puzzles have been included in interactive books, meaning that the direction of the narrative is, in part, dictated by the reader's answers to the problems. This allows for the possibility of the reader receiving immediate feedback on their answers, although this feedback may be explanatory, or simply corrective. Moreno (2004) used the Design-a-Plant microworld, mentioned earlier (Moreno and Mayer 2000), to test the *guided feedback hypothesis*: 'low prior-knowledge students learn best from discovery-based multimedia environments when pedagogical agents guide them with explanatory feedback'

(Moreno, 2004, p102). To test this, students were involved in the Design-a-Plant learning environment with the pedagogical agent providing one of two different types of feedback. In explanatory feedback (condition 1), students were provided with a verbal explanation as to why they were correct or not by the computer agent. In the corrective feedback (condition 2) however, students were simply told by the agent whether they were correct or not, without explanation. The results showed that the students under the explanatory feedback condition achieved 'better retention and transfer scores, and rate[d] the difficulty and helpfulness of the instructional program more favorably' (Moreno, 2005, p106). Consequently, explanatory feedback is preferable to corrective feedback in interactive texts.

Recommendation 4: Interactive narrative should incorporate explanatory, in preference to corrective, feedback.

Humanising science through history

Matthews (2015) has argued that humanizing science through the use of history makes the subject more attractive to many students. This has been backed up by empirical research; Klassen & Klassen (2014, p1503) cite a series of studies that demonstrate how 'historically based stories … promote the desired student interest and motivation by presenting humanistic episodes that explicitly include scientific content'.

Klassen (2007) answers the question of how to integrate history into a science story by arguing its use must be 'consistent with the historical record' (Klassen, 2007, p337). Naturally, this includes avoiding anachronisms and, although poetic license is an inevitable necessity, accuracy must still be an objective. However, both Clough (2011) and Klassen & Klassen (2014) come to adopt the stronger position that using a scientist's actual words adds authenticity as well as accentuating 'the human side of science' (Clough, 2011, p704). Consequently, primary sources should play a vital role in the writing of interactive texts for learning science. When this can't be achieved, any subsequently generated text attributed to a real-life character must be faithful to history and not contradict it.

Recommendation 5: Narrative should include a historical (or contemporary) setting and draw on primary sources from those settings as much as possible.

Elucidating their suggestion of using 'explanatory stories' in science education, Millar & Osborne (1998) felt it imperative that 'young people see these 'stories' not as 'given' knowledge but as the product of sustained inquiry by individuals working in social and historical contexts' (Miller & Osborne, 1998, p2014). Through deploying historical narratives which are as accurate as possible, the importance of sustained enquiry and working with others is likely to be foregrounded in science. However, this may still not be enough to counter the belief of many pupils that science is only for 'brainy' individuals (ASPIRES, 2013). We therefore recommend that the role of teamwork and building on the work of others is made explicit within historical (and contemporary) narratives:

Recommendation 6: Narrative should make it clear that scientists don't work in isolation and instead build on the work of others.

As we shall discuss shortly, this was particularly important in the text developed around the work of Isaac Newton.

Beware seductive details!

Despite the need for authentic narrative in science education, it has been shown that incorporating entertaining yet irrelevant information in expository text can have a detrimental effect on recall and problem-solving transfer (Harp & Mayer 1998). Often, these 'seductive details' prove more memorable than the material they were meant to add colour to. Harp and

Mayer investigated these ‘adjuncts’ and proposed that ‘in their search for ways to construct meaning from the text, readers are drawn to seductive details as an organizing context’ (Harp & Mayer, 1998, p432).

Lehman et al. (2007) extended the work of Harp and Mayer and examined the impact of seductive details inserted into a technical, scientific text. They suggest that the details reduced attention on principle information, disrupt the reader’s attempts to build causal chains (in working memory) and diverted readers through recall of inappropriate knowledge. However, they also noted that seductive details don’t always interfere with learning; not all seductive details are equal – how common such details are, how context dependent they are and the overall coherence of the text all played a role in determining whether learning was affected or not. As acknowledged by Lehman et al. (2007), this would explain earlier studies that found no evidence of the negative effects of seductive details when using short, technical texts (e.g. Garner & Gillingham 1991; Schraw 1998). Buselle & Bilandzic (2009) frame distractions that reduce narrative engagement as ‘the presence of thoughts that are unrelated to the narrative’ (Buselle & Bilandzic, 2009, p326). This is helpful in considering why irrelevant seductive details will interfere with the comprehension of a narrative but appropriate ones will not.

Recommendation 7: Narratives need to be as coherent as possible in order to steer clear of the possible dangers of distracting seductive details.

1.2 Developing an Interactive Fiction Book to Promote Understanding of Forces

Having established some of the benefits of using narrative to improve understanding within science education, we will here report on an attempt to use the vehicle of interactive fiction to improve readers' conceptual understanding of Newton's three laws of motion. The topic was chosen owing to the wealth of research on conceptual difficulties that students have in forces (e.g. Ioannides & Vosniadou, 2004; diSessa et al., 2002; diSessa et al. 2004), and the availability of an established tool for assessing these: the Force Concept Inventory (FCI) (Hestenes & Swackhamer, 1992); a thirty question multiple choice test. Furthermore, the richness of the historic context in and around forces informed the decision; few figures within the history of science have greater iconic status than Sir Isaac Newton, who has literally been placed on a pedestal throughout the western world. This status has the potential to distance students from Newton, a member of the pantheon of 'dead white male' scientists. This provides a further reason that considering Newton as a key figure within the developed interactive fiction would allow testing of the recommendations developed above, particularly those around personalization and historic context. We know that Newton suffered from the persistence of similar conceptual difficulties to those acknowledged by many school children (Steinberg et al. 1990), and that he acknowledged 'standing on the shoulders of giants', highlighting that his work was not conducted in isolation. A test of the developed book was how well these could be foregrounded in order to engage students in narrative.

Conceptual and contextual understanding of forces

A study by Maria and Johnson (1989) considered whether the genre of a text influenced its ability to correct misconceptions. Also called 'preconceptions' and 'alternative frameworks' (Treagust 1988), the term misconception is used to describe a belief that is at odds with that of the scientific community. This is an area that has led to a considerable amount of research, however there remains no single approach to understanding student conceptions and concept change in science (Amin et al. 2014). Forces remains a topic of particular interest to conceptual change researchers, and has formed the foundation for continuing debates around the nature of student conceptions (Ioannides & Vosniadou 2002; diSessa et al. 2004; Özdemir, & Clark 2007; Brown & Hammer 2008 and Özdemir, & Clark 2009). A prominent thread of these debates has been between those who suggest that young people have 'naïve theories' of the world which are nevertheless coherent in their composition, and those who see pupil understanding as more fragmented.

Ioannides & Vosniadou (2002), who might be classified as focusing on coherent theories, found that the vast majority of students' interpretations of force, whatever the context, fell into one of seven categories, and that there was a clear sense of progression in these categories. Each category is 'relatively well-defined and internally consistent' (Ioannides & Vosniadou, 2002, p1). However, when diSessa et al (2004) conducted an experiment intended to follow closely Ioannides & Vosniadou's (2002), they found an inconsistency in students' answers when questioned across a range of contexts e.g. the same student would use different meanings of the word 'force' depending on the question. Vosniadou and diSessa can be seen as representing two camps in the coherence vs fragmentation debate regarding naïve physical ideas.

Hugely relevant to this debate, and to this paper, is the extent to which context matters to an understanding of force. Özdemir and Clark (2009) ran a replication study and suggested that Ioannides' and Vosniadou's (2002) 'coding categories are so broad and soft that many students expressed multiple different meanings across contexts that still fell within the broad categories' (Özdemir and Clark, 2009, p594). This meant that the coherence of students' knowledge could be easily overestimated. Instead, they suggest that 'even small contextual variations may affect students' interpretations of force and, thus, this causes fragmentation in their causal responses' (Özdemir and Clark 2007, p357). This study, along with others, suggests the contextual sensitivity of students' conceptual understanding (Southerland at al 2001; diSessa et al. 2002; Thaden-Koach et al. 2006; Wagner 2006 and Clark 2006). In extending conceptual change research, further frames can be found to support for the context-specific nature of learning. For example, sociocultural views of conceptual change suggest that social context influences learning (Mercer, 2007). Furthermore, the field of *situated cognition* recognises the interaction of context and cognition in problem solving (see Kirsh, 2009). Whilst we are not able to engage with these fields fully here, one clear implication for the use of narrative in science education is that instruction should be context-driven and that the same phenomena should be presented in multiple ways. As Klassen makes clear, a historical science story can be considered 'a valuable… part of providing a rich and diversely connected context for student learning' (Klassen 2007, p335).

Recommendation 8: Narrative should utilise a range of contexts and settings in treatment of conceptual understanding.

The Force Concept Inventory

The well-established, Force Concept Inventory (FCI), provided a useful vehicle in both the development of an interactive narrative around the topic of forces, and its subsequent testing, because it is ready-made and its reliability has been demonstrated (Hake 1998, Henderson 2002; Lasry et al. 2011). In its current guise (version 2, 1995), the FCI is a 30 question multiple-choice test that assesses six 'conceptual dimensions' of student understanding of forces: Kinematics, Newton's First Law, Newton's Second Law, Newton's

Third Law, the Superposition Principle and Kinds of Force (Hestenes at el. 1992), although some of the questions relate to more than one. Hestenes & Halloun (1995) argue that there is compelling evidence that the total FCI score a student secures correlates strongly with what they term Newtonian thinking (understanding force and motion as described by Newton's three laws) and Newtonian skills, such as problem solving with reference to forces.

The authors are at pains to make it clear that the inventory's power rests in its completeness rather than with any individual question, as is supported by Lasry et al. (2011). This placed constraints on the required length of narrative to incorporate these questions, although this was to some extent ameliorated by the range of contexts presented within the inventory, allowing a richer narrative.

To make the narrative side of the developed interactive fiction as strong as possible, it was necessary for the narrative to take priority and the FCI to feed the questions into this narrative rather than the narrative have to be constructed around the FCI. Significantly, this meant incorporating the FCI's questions in an order that is different to the original's. This raises the issue of whether this affects the tool's reliability. Lasry et al.'s (2011) study of FCI reliability determined that it was 'internally consistent' and that 'scores on different halves of the FCI are highly correlated' (Lasry et al., 2011, p910). This conclusion was produced using the Kuder & Richardson's (1937) reliability coefficient, KR-20. Crucially, this method 'estimates the average correlation between *all possible halves*' (Lasry et al., 2011, p909). For example, a 30 question test could be split into halves in many ways, such as the first fifteen and second fifteen questions, odd and even numbered questions etc. Kuder and Richardson came up with a method for determining the average correlation between two halves of a test across all possible ways of splitting it into two halves. A value of '0' means there is no internal consistency and '1' shows perfect consistency. The KR-20 value of the FCI test was 0.900. This suggests that reordering the questions is unlikely to affect the reliability of the FCI. Furthermore, it allows the possibility of assessing whether a reader's performance changes during the reading of the book by comparing students' correct answer scores between the first and second half of the questions and, if it does, whether the effect is a positive or negative one. This became significant in the testing of the interactive fiction developed (as is detailed below).

A Forceful Adventure 1.0

Having reviewed literature on the use of narrative, with a particular focus upon the use of interactive fiction to promote conceptual change in science education, we arrived at eight recommendations for the book to be developed and tested:

Recommendations for the use of Narrative in Science Education:

1. Narrative should incorporate Avraamidou and Osborne's (2009) necessary components: purpose; events; structure; time; agency; narrator; reader.

2. Narrative should employ the personalisation principle and pedagogical agents whenever possible.

3. Interactive narrative texts should be written in the second person present, in line with the commercially accepted format.

4. Interactive narrative should incorporate explanatory, in preference to corrective, feedback.

5. Narrative should include a historical (or contemporary) setting and draw on primary sources from those settings as much as possible.

6. Narrative should make it clear that scientists don't work in isolation and instead build on the work of others.

7. Narratives need to be as coherent as possible in order to steer clear of the possible dangers of distracting seductive details.

8. Narrative should utilise a range of contexts and settings in treatment of conceptual understanding.

These recommendations were utilized in order to develop and then test a 131 section interactive work of fiction: *A Forceful Adventure 1.0,* centred on the school topic of Forces. Set in two temporal settings, a contemporary school environment and 1680s Cambridge, it involves answering the Force Concept Inventory's multiple-choice questions. The reader's answers to these determine which section of text they then turn to next. Below is a sample section:

73

Deep in thought, you only now realise you've left Cambridge and are walking along a river path with bright wildflowers dotting the adjacent fields like stars in the night sky. Newton, in his fading orange greatcoat, is tossing a metal ball into the air as he moves while holding his walking stick in the other hand. You put your homework back in your bag.

'You mentioned Galileo last time we met,' he says. 'One of his friends, Stelluti, gives a lovely account of a brilliant experiment Galileo performed. Stelluti describes how he was on a six-oar boat with Galileo that was going very fast across Lake Piediluco. Galileo leaned across and asked him if had anything heavy on him. Stelluti replied that he had the key to his room and handed it to Galileo. With the boat still moving rapidly, Galileo threw the key high into the air. Stelluti gave a sharp intake of breath and waited for the key to drop into the lake and be lost forever. To his great surprise, the key fell down between them despite the boat having moved some eight to ten arm lengths in the time the key was in the air. This was because, besides going up, the key had also acquired forward velocity from the motion of the boat.'

Newton stops walking but continues to throw the metal ball straight up.

‘Consider the motion of the ball after it has left my hand but before I touch it again. Assume that the forces exerted by the air are negligible. What would you say the force(s) acting on the ball is (are):

(A) A downward force of gravity along with a steadily decreasing upward force. Turn to **77**.

(B) A steadily decreasing upward force from the moment it leaves my hand until it reaches its highest point; on the way down there is a steadily increasing downward force of gravity as the object gets closer to the earth. Turn to **107**.

(C) An almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down there is only a constant downward force of gravity. Turn to **119**.

(D) An almost constant downward force of gravity only. Turn to **58**.

(E) None of the above. The ball falls back to my hand because of its natural tendency to rest on the surface of the earth. Turn to **104**.

We will now illustrate how the recommendations above were incorporated into the developed text.

Incorporation of the necessary components of narrative and writing in the second person (Recommendation 1)

A Forceful Adventure opens with the following:

“Driiiiiiiiiiinnnnnnnnnggggggggg!’

The nine o'clock bell announces the start of yet another day at school. You lug your bulging bag out of form time and head towards the science labs.

The text is written in the second person and there is a sense of agency (you and others are experiencing events). It is clear from the opening section that the text has a purpose - there's an anticipated understanding of an aspect of the natural world as shown by the question the class is intended to answer:

'Right, everyone, we're starting a brand new topic today – forces,' says Ms Sharp. 'By the end of the topic, I want us to be able to begin to answer the question, "Why do objects – a ball, a bullet, the moon – move the way they do?"

There is a connected sequence of events (a question is asked which prompts a response); a structure for the book has been immediately set up, one that begins unfolding immediately. There is also an implicit narrator addressing the reader who would undoubtedly recognise this text as a narrative.

Avraamidou and Osborne (2009) also identified time, where narratives occupy the past, as a key component. This was also the view of Norris et al (2005) and (Klassen & Klassen 2014). However, Chatman (1978), an influence on Avraamidou and Osborne's list of necessary components, acknowledged that narrative could be written in the present tense.

The extracts above make clear that what is to come is an understanding of forces and Ms Sharp's next question establishes that the class will be studying Newtonian forces:

So, can anyone think what Sir Isaac Newton has to do with forces?'

There's a pause and then a few tentative arms begin to rise. Then a couple more. Yours most definitely stays fixed to the desk.'

During the course of the narrative Newton's three laws of motion are stated, explained, reinforced and tested. The explanations come from a combination of Ms Sharp (the class teacher) and Newton, who both act as pedagogical agents in the text. Reinforcement comes from multiple examples of the laws in action along with associated explanation from Ms Sharp and Newton. The testing takes place through the use of the FCI's questions.

Employing the personalisation principle and pedagogical agents (Recommendation 2) and writing in the second person present (Recommendation 3)

As the entire book is written in the second person, the personalisation principle is an inherent feature of the text. For example, here is Newton responding to an incorrect answer in which he's also acting as a pedagogical agent:

'It's true that objects would eventually reach a constant, or terminal, speed but this takes time. Would your answer be different if I said it was you who was dropped from my window? Do you think you'd reach a constant speed before you hit the ground? Whatever you say about the falling stone should also apply to yourself – we're only changing the mass.' (from section 11)

It should be noted that the use of the personalisation principle and a pedagogic agent doesn't necessarily mean that a text is a narrative. Similarly, interactive fiction need not necessarily be in the second person. We have thus kept recommendations 2 and 3 separate, although in practice they are both easily satisfied by writing in the second person.

Providing explanatory, as opposed to corrective, feedback (Recommendation 4)

Explanatory as opposed to corrective feedback was provided in about two-thirds of the FCI questions. This highlighted difficulties in deploying all 30 FCI questions in that it was a time consuming process, but also lengthened the text considerably, reducing its usability. A balance between these factors was sought. Some of the explanations also made connections between different contexts:

'After selecting (B) it suddenly dawns on you that the car/truck question is just like the horse and cart problem you discussed with Newton. In that instance, the horse was pulling the cart and here the small car is pushing the truck. The principle is still the same…' (from section 2).

In addition to explanatory feedback, the book also contained exposition, delivered either by the teacher or Newton. This was written in a conversational style and often involved interaction between Ms Sharp and other members of the class.

The decision to use all of the FCI's questions placed considerable constraints on the writing of the tool. The most significant impact was the final length of the resource. Even with shortcuts taken, which will be explained shortly, the final length of the text proved to be just over 25,000 words. While no reader would read every word given the non-linear choice-based nature of the book, it is still amounts to a large amount of reading and considerably more than any of the narrative examples featured in the literature review.

In one part of the book, the protagonist has to complete homework comprising ten of the 30 FCI questions. It is possible to ask Newton for help with each question but they are asked one after the other and, for these questions, corrective feedback only is provided a few sections later when you receive the correct answers from the teacher. So, ten out of the 30 questions are asked in a more conventional manner. This was less than ideal but necessary if a full first draft was to be completed and tested in the time available.

Given the need to fit the questions into the narrative, it was occasionally necessary to change their wording. This was kept to a minimum and only done to improve the flow of the text (an added benefit was that it often ended up employing the personalisation principle). Here is an example:

(Original FCI text)
‘Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a single story building at the same instant of time. The time it takes the balls to reach the ground below will be:’

to

(A Forceful Adventure text)
‘Take these two metal balls. They are the same size but one weighs twice as much as the other. If you were to drop the balls from the top of a single-storey building at the same instant, how long would it take the balls to reach the ground below?’ (from section 1)

Having a historical setting and drawing on primary sources as much as possible (Recommendation 5)

The story has two temporal settings: modern day and Cambridge in 1685. The former allows for conventional lessons to take place within a context that allows other agents to facilitate learning (e.g. the teacher and other students) as well as providing an imperative for learning about forces. Cambridge in 1685 allows greater colour and flexibility as well as the opportunity to interact with a science ‘great’, Isaac Newton, who acts as a guide through parts of the story.

The school setting is intended to be both familiar and generic – there is very little sense of the environment other than key semiotic signs such as a school bell, a teacher, other students and a whiteboard among other things. Cambridge, however, contains much more description, which is intended to be as accurate as possible. Preparation for writing the story included a visit to Trinity College as well as a variety of reading regarding Newton's life and his time at the college in particular. The location of Newton's rooms, garden and lab have been drawn from that reading (e.g. Adrian 1963, Westfall 1983 and Spargo 2005). For example, the parts underlined in the following section are all based on this reading.

'As you step out onto the grass, the southern sun blinds you temporarily. The beauty of the garden, with not a single weed in sight, is offset somewhat by an acrid smell *in the air. To your left is the exterior of the college chapel with three bays formed by the buttresses of the chapel. In the southeastern bay is a wooden structure with smoke streaming from a brick chimney at one end of the shed.' (from section 110)*

The use of primary or secondary sources extended to physical descriptions where possible. For example, the following description of Newton:

'a relatively short middle-aged man with long, thick grey hair that looks like it hasn't seen a comb in quite a while, ambling through the gatehouse in a distracted manner.' (from the 'Background')

was based on the contents of a letter by Newton's assistant during the 1680s, Humphrey Newton (Newton 1728).

In keeping with Klassen (2007), Newton's actual words were used whenever possible e.g. one version of his 1st law is taken directly from one of his notebooks (Newton 2013). I've also relied on the first English translation of the Principia.

You glance at the sheet on the top and it dawns on you that these are Newton's notes. Without pausing, you begin to read aloud what he's written:

'A quantity will always move on in the same streight line (not changing the determination nor celerity of its motion) unlesse some externall cause divert it.' (from section 1)

However, despite the existence of letters, notebooks and records containing millions of words written by Newton, there are no personal diaries or correspondence containing comments regarding his personal or private thoughts. This reduced the ability to show Newton's character through his own words.

I felt it important to avoid any historical inaccuracies such as Newton using the formula F = ma and the difference between the laws as stated by Newton and Ms Sharp is intended to illustrate the refinement of science and its expression over time.

Conveying that scientists build on the work of others (Recommendation 6)

The fact Newton looked at the work of other scientists is principally demonstrated through his discussion and exposition of parts of Galileo's work as well his explanation that motivation for writing the Principia came from a conversation with Edmond Halley. Nevertheless, a significant problem in writing the book was Newton's character. For all his brilliance and hard work, he was an introverted, secretive, suspicious and insecure person prone to violent outbursts (Westfall 1983; The Newton Project). While the book is as faithful as possible to the words and circumstances surrounding Newton's work on Forces, the character portrayed in the book doesn't reflect the man's complexity. In an attempt to address this, the very end of the book, after the adventure is over, contains the excellent text on 'Isaac Newton's Personal Life' from The Newton Project based at the University of

Sussex, UK.

Avoiding seductive details (Recommendation 7)

As noted above, including the entire set of 30 FCI questions limited the narrative freedom to some extent. The temptation to include further detail was therefore tempered by the need for brevity. However, in making the contexts authentic (relating to Recommendations 5 and 8) and wanting to ensure that the texts were engaging for readers, details were included that were directly relevant to the context or problems presented.

To exemplify this, consider the below passage from the book:

‘Stand the spring on the table and squash it. Keep it squashed – what do you notice?’

‘I feel the spring pressing back,’ you say.

‘That’s right. A force is one side of an interaction; the interaction takes place between two bodies, working equally strongly in the two opposite directions. In this case the spring is one body and your hand is the other. The same is happening between the table and the spring.’

You walk to his window and look out onto Newton’s garden and Trinity Street beyond. You notice a horse and cart just like Ms Sharp mentioned. A thick-set man appears to be berating the horse. It clearly doesn’t want to pull the cart, which isn’t a surprise given how much furniture is piled on top of it. A thought comes to you and you turn back to Newton.

‘How does it work with that horse and cart?’ you ask, drawing Newton’s attention to the commotion outside.

'What do you mean?'

'Well, isn't the horse right to refuse its owner's demands? Doesn't your law say that the cart is just going to pull back on the horse with an equal force? Doesn't that mean they would remain stationary? I mean I obviously know that the horse is capable of pulling the cart forward.'

Just as you say that, the horse gives out a large snort and begins to pull the cart forwards.

Here, narrative engagement is balanced with a focus upon the problem being considered. The reference to Newton's garden situates the narrative historically, and the dispute between the man and the horse aims at engagement whilst not deviating far from the problem of the spring being reflected upon. Ms Sharp having referred to the earlier horse also creates narrative continuity. In the above example, the spring is presented as just that, rather than through the use of a toy or other narrative device which would provide seductive detail. Similarly, it might be tempting to further detail the relationship between the man and the horse, or explore the role of horses in Newton's Cambridge. Again, the focus is on providing enough for a compelling narrative, without overcomplicating the examples through which the learning takes place in the text.

Utilising a range of contexts and settings in treatment of forces (Recommendation 8)

There are sixteen different contexts in the FCI to which A Forceful Adventure 1.0 adds a further ten.

These are the contexts featured in the Force Concept Inventory:

- Two metal balls of the same size but with one twice the weight of the other being dropped.
- A stone being dropped.
- A large truck interacting with a small compact car.
- A ball shot along a horizontal part-circular frictionless channel.
- A steel ball attached to a string being swung in a circular path in a horizontal plane.
- A hockey puck sliding on a frictional surface.
- A ball being fired by a cannon from the top of a cliff.
- A bowling ball falling out of the cargo bay of an airliner flying in a horizontal direction.
- An elevator being lifted up an elevator shaft by a steel cable.
- A boy swinging on a rope.
- The motion of two blocks.
- A rocket in outer space subject to no outside forces.
- A woman pushing a box.
- Two students of different masses on office chairs with castors with one student pushing off the other.
- An empty office chair on at rest on the floor.
- A tennis ball moving against a strong wind.

In addition, the following contexts are added via the narrative:

- The European Space Station orbiting the Earth, along with the astronauts inside it.
- A horse pulling a cart.
- A stone being pressed.
- A spring being pressed.
- Throwing a key vertically into the air while in a moving boat.
- A rollercoaster taking a bend in the track.
- A car going round a corner.
- Jumping off a boat or dinghy.
- Firing a gun.
- Weight on the Earth and Moon.

2. Testing A Forceful Adventure 1.0

As well as developing recommendations for the development of interactive fiction within science education, this paper reports on the initial testing of the developed example: *A Forceful Adventure 1.0*. In this initial, exploratory stage, the primary intention was to test whether the book had an effect upon the conceptual understanding of students who engaged with it. We also considered whether the book had a differing effect upon the conceptual understanding of students who study physics at a higher level, from those who do not. A secondary focus was how far the students enjoyed engaging with the text and whether it influenced their confidence in relation to the topic of forces. This secondary focus is born of a desire for the book to have pedagogical utility in that students would use it, but also to ensure that the compromises discussed earlier in terms of narrative do not negatively affect enjoyment and subsequent utility of such a text. We furthermore wanted to find out if pupils who read for pleasure more than others were more likely to have a positive response to the book. The below research questions were thus explored:

1. Does the interactive fiction improve the ability of readers to answer the Force Concept Inventory questions embedded within the narrative?
2. Is there a differing effect size for students who study physics to an advanced level, compared to those who do not?
3. Do readers report that engaging with the book is enjoyable?
4. Does reported enjoyment differ for those who regularly read for pleasure, and those who do not?
5. Does reading the book change reported confidence in the topic of forces?

Sample and Methods

Given the FCI was originally developed to assess college students' Newtonian thinking (Hestenes et al. 1992), the population chosen for the study was students in three Year 12 chemistry classes in London, UK (pupils aged 16 and 17). Across the three classes, 27 students took part (16 female, 11 male). Eight had studied physics during the year and the remainder only up to the age of 16 (the end of compulsory science education in the UK). Students were given up to two hours to engage with the book in silence, with a 5 minute break (during which no mention of the task was allowed).

In order to investigate the first two research questions, we made use of Lasry et al.'s (2011) finding that the Force Concept Inventory is 'internally consistent' (KR-20 = 0.900) and that 'scores on different halves of the FCI are highly correlated' (Lasry et al., 2011, p910). As discussed earlier, their method estimated the average correlation between *all possible halves* of the FCI. This means that a measure of improvement across the narrative fiction can be gained by comparing the score on the 15 FCI questions first encountered within the the developed book, and the last 15 FCI questions encountered during reading.
A tracking sheet was used as a way for each reader to record their route through the adventure along with their answers to certain questions in the text. This provided a way of checking a student's answers for every FCI question, revealing which questions were answered correctly first time. This also allowed tracking of the route that each pupil took through the text, allowing the scoring of the FCI questions, well as the possibility of breaking down the FCI questions into two halves for comparison of progress in understanding.

In order to answer research questions 3, 4, and 5, relating to enjoyment and confidence, questionnaires were used both pre and post engagement with the interactive fiction. Questionnaires were designed taking into account Bell's (2007) suggestions that questions are short and employ a straight-forward syntax, avoiding negative formulations. Likert-type scales were used in a number of questions and were verbally labelled. Pre-questionnaires were used to measure confidence on a five-point verbal scale (very poor, poor, okay, good, very good) in the topic of Forces. This was repeated in the post-reading questionnaire so that any perception of change could be ascertained, in line with other studies into confidence in science education (e.g. Mulop et al 2014; Shoemaker 2010). A question was also included in the pre-questionnaire asking pupils how many books are read for pleasure each month, in order to determine if there was a correlation between this figure and their rated enjoyment of the text (reported in the post-questionnaire). It should be noted from the outset that the measures of enjoyment and confidence are based on a single question each, so provide no more than an initial indication of these constructs. Pupils were furthermore asked to rate the explanations and puzzles in A Forceful Adventure 1.0, to provide three words to describe what they thought of the text, and to answer open ended questions asking for suggestions on how these could be improved. The responses to these questions were intended to provide feedback on the text itself, and as such are not reported in detail in this paper.

Analysis

All 27 participants provided useable data in the questionnaires, and 23 tracking sheets were considered after excluding those with missing information. Likert-type scale questions were coded numerically, and the median and mode were used as the measure of central tendency for this ordinal data (Jamieson 2004). It has been established that ordinal data can be used in parametric tests such as *t* tests and Pearson correlations (Norman 2010; Sullivan & Artino 2013) and that parametric tests tend to be generally robust, often more so than nonparametric tests (Norman 2010). As such, paired Sample T-tests were undertaken in order to determine the statistical significance (the *P* value) of any results regarding differences between groups. Following Ghasemi & Zahediasl (2012), a Q-Q plot and Shapiro-Wilk test were used to confirm that the dependent variable (the 2nd half scores) did not deviate significantly from a normal distribution ($p=0.56 > 0.05$). Sullivan and Feinn (2012) have argued that statistical significance isn't enough and that effect size (d) must also be taken into account. As such, effect size was calculated using first 15 FCI questions encountered in the book as the ‘control’ condition and the second 15 encountered were treated as the ‘experimental’ condition. This is in line with Lasry et al.’s (2011) finding that scores on all possible halves of the FCI are highly correlated (as discussed earlier). The standard deviation used was a ‘pooled’ value from both groups, and the margin of error in estimating the effect sizes was calculated based on a 95% confidence interval.

3. Results

Effect sizes

An analysis of the tracking sheets showed that readers of *A Forceful Adventure 1.0* got better at answering the FCI’s questions when comparing their success in the first 15 FCI items encountered (Mean=50.78%, Standard deviation=25.159%) with the second 15 items encountered (M=66.87%, sd=19.652%). This difference, 16.087%, was significant $t(22)$ = 4.243, p= .000, and represented a medium-sized effect, d = 0.71 with a 95% confidence interval [0.11, 1.31].

It was unclear whether readers of *A Forceful Adventure 1.0* who were also studying physics got better at answering the FCI’s questions when comparing their success in the first 15 FCI items within the book (M=75.57%, SE=14.010%) with the second 15 FCI items (M=82.29%, SE=12.325%). This difference, 6.714%, was not significant $t(6)$ = 1.426, p= .204, and represented a medium-sized effect, d = 0.51 with a 95% confidence interval [-0.52, 1.60].

Readers of *A Forceful Adventure 1.0* who had studied Physics only up to the age of 16, got better at answering the FCI's questions when comparing their success in the first 15 FCI items in the book (M=39.94%, SE=20.92%) with the second 15 items (M=60.13%, SE=18.608%). This difference, 20.188%, was significant $t(15) = 4.233$, p= .001, and represented a large-sized effect, d = 1.02 with a 95% confidence interval [0.28, 1.77].

Enjoyment

The distribution of student ratings of their enjoyment of engaging with *A Forceful Adventure 1.0* was skewed towards the positive, but with both the median and the mode were an enjoyment rating of 'okay' (see Figure 1).

Figure 1: Enjoyment ratings

A Pearson Correlation was calculated to ascertain whether there was a correlation between the number of books read for pleasure and the enjoyment of *A Forceful Adventure 1.0*. There was a very small correlation (0.125) with no statistical significance ($p > 0.05$). This was investigated further by seeing if there was a connection between reading for pleasure (students who read any number of books each month for pleasure) and enjoyment. Again, there was a very small correlation (0.177) with no statistical significance ($p > 0.05$). These results suggest that the enjoyment of the book was not affected by the reading habits of individuals. In addition, there was no connection between reading for pleasure and how many FCI questions pupils got correct ($r = 0.041$, $p > 0.05$).

Confidence

Figure 2: Confidence rating pre- and post-reading

The median and mode of participants' confidence was 3 ('okay') and post-reading it was 4 ('good') (see Figure 2). However, on average, there was no significant change in a reader's confidence in forces after reading *A Forceful Adventure 1.0*. This also proved the case when the data was broken down into those students who had studied physics at a higher level compared to those who stopped at the age of 16. While it has been shown that confidence can be used as a measure of learning (Mulop et al 2014; Shoemaker 2010), these results can still be considered difficult to unpick as the feedback inherent in the book provides the reader with a way of better assessing their understanding of forces; a decrease in confidence may come about because they overestimated their understanding of forces and an increase can occur because they underestimated or because they felt themselves improve during the reading. For example, Lichtenstein and Fischoff (1977, p159) asked the question 'do those who know more also know more about how much they know?' and found that people are prone to systematic biases, most commonly over confidence.

4. Discussion

A Forceful Adventure has a sizeable, positive effect on the average student's ability to answer Force Concept Inventory questions. The effect size was moderate to large (d=0.71). This effect is even more pronounced when the data is broken down and students who studied physics only up to the age of 16 are examined. We considered this relevant as these students will have only encountered Newton's laws in their studies once before, so may have naïve understandings of force. Here the effect is large (d=1.02). The results of students who were also studying further physics are less certain – there is a medium sized effect and the result isn't statistically significant (p=.204). This may well be because the sample size is too small, but could of course be due to the more secure understanding of forces amongst these students, who have formally studied mechanics both in their compulsory science education, and post-compulsory studies in physics. The results allow us to conclude that the developed interactive fiction, increases a reader's ability to answer FCI questions or, possibly, be better at knowing what the wrong answer is. However, Hestenes et al. (1992, p.2) have argued that the FCI is a 'very good detector of Newtonian thinking' and Lasry et al. (2011, p911) confirmed that 'the total FCI score measure a unique construct'.

To contextualise the effect sizes seen within this study, consider Hake's (2002) report on his earlier, seminal meta-study of more than six-thousand undergraduate students sitting introductory mechanics courses in the USA (Hake, 1998). For the studies from which an effect size could be derived, he reports that students who sat nine 'traditional', lecture-based undergraduate mechanics courses (1620 students) showed an average effect size of d=0.88 across pre- and post-testing with the force concept inventory. For students who were taught undergraduate mechanics through 25 'Interactive Engagement' courses (1843 students), involving tutor interaction and immediate feedback, a large effect size of d=2.16 was seen. In this context, the overall effect size for all students reading *A Forceful Adventure* (d=0.71), and the effect size for those not studying further physics (d=1.02) are comparable to an entire traditional lecture course on mechanics, but not as large as seen through interactive teaching over a course. These comparisons are very encouraging given that the effect size here is seen from a single, self-directed reading of an interactive fiction book.

Of course, caution should be expressed in that the sample within the present study was very small, and the assessment and 'intervention' (the interactive fiction itself) were both contained within the book. This means that the findings are not directly comparable to pre- and post- testing using the Force Concept Inventory. As discussed earlier however, we could not find any existing research into the use of interactive fiction books as a mode of improving conceptual understanding in science education. Whilst treating the findings here with caution therefore, we encourage further research into the apparent potential of interactive fiction books in science.

In order for interactive fiction books to have pedagogical merit, they must also be enjoyed by those who engage with them. This was therefore a secondary focus of this small-scale study. It is clear from the results that a participant's reading habits are unlikely to influence their enjoyment of the book. Although the median of enjoyment rating was only 'okay', 'interesting' was used by ten of the 27 students to describe their feelings towards *A Forceful Adventure 1.0*. A further eight used at least one positive word coded as 'engagement/enjoyment'. It should be recalled however that these findings are based on a single rating and further open response question. One relevant finding that came out of the various studies on seductive details was that interest played a role in the learning process (Schraw & Lehman 2001). As we will come to shortly though, the limitations upon the narrative style and the length of the book, both necessitated by the inclusion of the entire Force Concept Inventory, may have prevented greater enjoyment of the book.

It is uncertain what conclusions can be drawn from the measurement of self-reported confidence pre and post-reading. Students who were also studying advanced level physics on average rated their confidence in the topic of Forces higher than those who had studied it only up to the age of 16. While the analysis shows a positive effect using the tool when comparing the first and second halves of the FCI questions, this may not be so obvious to the reader. Equally, it may be that they'd simply overestimated their understanding in the first place. The Median did increase from 3 to 4 but there was not statistical significance regarding confidence before and after, even when the analysis was broken down into those studying further physics, and those not.

Further developing interactive fiction

The empirical study presented supports the recommendations developed within the literature review above, by showing that a text which incorporated the recommendations had a moderate to large effect upon conceptual understanding of a group of students. The recommendations draw on studies which isolate and test the use of principles such as personalisation, corrective feedback and use of context. Here we have brought these together in a single text and, recognising the limitations of the small-scale study, have presented empirical support for this. We therefore suggest that these recommendations are incorporated into further texts in exploring the potential of interactive fiction to support learning in science. Of course, each interactive text must be developed through consideration of exactly *how* these recommendations can be incorporated. This requires stylistic decisions and skill as an author, as well as an understanding of the context being described through narrative.

We discussed earlier the compromises in ensuring that the narrative flowed, despite including the necessary aspects of the Force Concept Inventory. This highlights the need to balance engaging narrative and reader enjoyment, with the desire to address significant topics of science. In the study reported here, the number and style of the questions within the FCI, meant that the length of the narrative and also some of its content were restricted. For example, Newton's character was not fully developed, which may have prevented the 'humanizing' of science that Matthews (2015) and Klassen & Klassen (2014) suggest. The use of the FCI to provide data on effectiveness of the book need not be replicated in other interactive fiction however, and authors may thus be freed up to develop more engaging and authentic narratives.

Moving away from established research tools, such as the FCI, would require the development of a rigorous frame for evaluating the effectiveness of developed texts. Such a frame would also need to consider the relative effectiveness of such texts across age ranges and levels of study. In this respect, Busselle & Bilandzic's *Narrative Engagement Scale* (2009) may be a good starting point, which 'distinguishes among four dimensions of experiential engagement in narratives: narrative understanding, attentional focus, emotional engagement, and narrative presence' (Busselle & Bilandzic, 2009, p321). However, a further possibility for incorporating established tools such as the FCI in interactive fiction is the separation of the narratives to fit across sequences of teaching. Around the world, the Force Concept Inventory (FCI), is typically used as way of measuring the effectiveness of the teaching of mechanics (Hake 1998; Savinainen & Scott 2002; Henderson 2002). This is achieved by comparing scores on the FCI test pre- and post-instruction. Henderson (2002) investigated whether 'giving the FCI as a pre-test influence[d] the post-test results' (Henderson, 2002, p543) and found there to be no effect.

A Forceful Adventure effectively had all three parts (pre-test, intervention, post-test) compressed together. Within a teaching sequence, interactive fiction could be applied to different purposes, perhaps most naturally as the instruction/intervention, with established tools forming the pre-test and post-tests. Narratives could however, be deployed as assessments of understanding before and after periods of instruction. If interactive fiction were to be utilised as a form of instruction (only), then explanatory feedback would take on increased importance, which may lengthen the text. However, without the requirement of testing, the text need not be read in one sitting and could be engaged with over a period of time. This might be done in tandem with other modes of instruction.

Concluding remarks

There is a wealth of research upon the use of narrative in learning which has only recently begun to find its way into consideration of science education. However, to date there has been little research on the use of interactive fiction, and books in particular within science education. The review of this literature within this paper has developed eight recommendations pertinent to the deployment of narrative within science education, with particular reference to interactive fiction and its potential in promoting conceptual change. It is hoped that these recommendations might form the basis of further study and testing of this potential.

The small scale study which utilised the recommendations to develop an interactive fiction book around Isaac Newton and his laws, has provided preliminary feedback upon the challenges and compromises inherent in such development. It has also provided initial evidence, albeit on a small scale, that interactive fiction of this type can have a positive effect upon the conceptual understanding of students in science. We hope this encouragement will promote further study into the role of narrative and its affective dimensions within science education.

References

Adrian, T. L. (1963). Newton's rooms in Trinity. *Notes and Records of the Royal Society of London*, *18*(1), 17-24.

Allchin, D. (2014, July). The episodic historical narrative as a structure to guide inquiry in science and nature of science education. In *10th international conference on the history of science and science education, University of Minnesota, Minneapolis, Minnesota*.

Amin, T.G., Smith, C.L. & Wiser, M. (2014). Student conceptions and conceptual change. In Abell, S. K. & Lederman, N. G. (Eds.), *Handbook of research on science education: Volume II*, 57-81, New York: Routledge.

Arya, D. J., & Maul, A. (2012). The role of the scientific discovery narrative in middle school science education: An experimental study. *Journal of Educational Psychology*, *104*(4), 1022-1032.

ASPIRES (2013) Young people’s science and career aspirations, age 10 –14. London: King’s College.

Auble, P. M., & Franks, J. J. (1978). The effects of effort toward comprehension on recall. *Memory & Cognition*, *6*, 20-25.

Avraamidou, L. & Osborne, J. (2009), The role of narrative in communicating science. *International Journal of Science Education*, *31*(12), 1683-1707

Bartlett, F. C. (1932/1995). *Remembering*. Cambridge, UK: Cambridge University Press.

Bell, A. (2007). Designing and testing questionnaires for children. *Journal of Research in Nursing*, *12*(5), 461-469.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. Reading Psychology, 29, 137-164.

Britton, B. K., Graesser, A. C., Glynn, S. M., Hamilton, T., & Penland, M. (1983). *Discourse Processes*, *6*, 39–57.

Brown, D. E. & Hammer, D. (2008). Conceptual change in physics. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change*, 127-154. New York: Routledge.

Bruner, J. (1986), *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, *18*(1), 1-21.

Bruner, J. (1996), *The Culture of Education*. Cambridge, MA: Harvard University Press.

Bruner, J. (1997). Celebrating divergence: Piaget and Vygotsky. *Human Development*, *40*(2), 63-73.

Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology*, *12*(4), 321-347.

Chatman, S. (1978). *Story and discourse: Narrative structure in fiction and film*. Ithaca, NY: Cornell University.

Clark, C., & Rumbold, K. (2006). Reading for pleasure: a research overview. London: National Literacy Trust.

Clark, R. C., & Mayer, R. E. (2011). *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning* (3rd Edition) San Francisco, CA: John Wiley & Sons.

Clough, M. P. (2011). The story behind the science: Bringing science and scientists to life in post-secondary science education. *Science & Education*, *20*(7-8), 701-717.

Crowder, N. A. (1963). On the differences between linear and intrinsic programing. *The Phi Delta Kappan*, *44*(6), 250- 254.

diSessa, A.A., Elby, A., & Hammer, D. (2002). J's epistemological stance and strategies. In G. Sinatra and Pintrich (Eds.), *Intentional conceptual change* (238-290). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Duit, R. (2009). Bibliography STCSE: Students' and teachers' conceptions and science education. *Kiel, Germany: University of Kiel.*

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*, 202-224.

Englert, C. S., & Hiebert, E. H. (1984). Children's sensitivity to expository text structure. *Journal of Educational Psychology*, *76*, 65–74.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance--A Critical Literature Review*. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.

Garner, R., & Gillingham, M. G. (1991). Topic knowledge, cognitive interest, and text recall: A microanalysis. The Journal of Experimental Education, 59(4), 310-319.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2), 486-9.

Glaser, B. G., Strauss, A. L., & Strutzel, E. (1968). The Discovery of Grounded Theory; Strategies for Qualitative Research. *Nursing Research*, *17*(4), 364

Graesser, A. C., Singer, M., Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. *Psychological Review*, *101*, 371–395.

Green, J. (2014), *You are the hero!* London: Snowbooks.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, *66*(1), 64-74.

Hake, R.R. 2002. "Lessons from the physics education reform effort," Ecology and Society 2: 28; online at < http://www.ecologyandsociety.org/vol5/iss2/art28/ >.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of educational psychology*, 90(3), 414

Henderson, C. (2002). Common concerns about the force concept inventory. *The Physics Teacher*, *40*(9), 542-547.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, *30*(3), 141-158. (Revised August 1995 by Ibrahim Halloun, Richard Hake, and Eugene Mosca)

Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory. *The Physics Teacher*, *33*(8), 502-506.

Hidi, S., Baird, W., & Hildyard, A. (1982). That's important but is it interesting? Two factors in text processing. *Advances in Psychology*, *8*, 63-75.

Ioannides, C., & Vosniadou, S. (2002). The changing meanings of force. *Cognitive Science Quarterly*, *2*(1), 5-62.

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical education*, *38*(12), 1217-1218.

Jetton, T. L. (1994). Information-driven versus story-driven: What children remember when they are reading informational stories. *Reading Psychology*, *15*(2), 109–130.

Katz, D (2016). TutorText: Doubleday series. Demian's gamebook web page http://www.gamebooks.org/show_series.php?id=457 [Accessed on 28th July 2016].

Keenan, J.M., Baillet, S.D., & Brown, P. (1987). The effect of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, *23*, 115–126.

Kirsh, D. (2009). Problem Solving and Situated Cognition. In Robbins, P. & Aydede, M. (eds.) *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press. pp264-306.

Kinzer, C. K., Hoffman, D. L., Turkay, S., Gunbas, N., & Chantes, P. (2011). Exploring motivation and comprehension of a narrative in a video game, book, and comic book format. *60th Yearbook of the Literacy Research Association*, 263-278.

Klassen, S. (2007). The application of historical narrative in science learning: The Atlantic cable story. *Science & Education*, *16*(3-5), 335-352.

Klassen, S., & Klassen, C. F. (2014). Science teaching with historically based stories: Theoretical and practical perspectives. In *International handbook of research in history, philosophy and science teaching* (pp. 1503-1529). Springer Netherlands.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2(3), 151–160.

Lancy, D. F., & Hayes, B. L. (1988). Interactive fiction and the reluctant reader. *The English Journal*, *77*(7), 42-46.

Lasry, N., Rosenfield, S., Dedic, H., Dahan, A., & Reshef, O. (2011). The puzzling reliability of the Force Concept Inventory. *American Journal of Physics*, *79*(9), 909-912.

Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology*, *32*(4), 569-587.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648.

Lester, J. C., Stone, B. A., & Stelling, G. D. (1999). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. In

Computational Models of Mixed-Initiative Interaction (pp. 185-228). Springer Netherlands.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know?. *Organizational behavior and human performance*, *20*(2), 159-183.

Lipson, S. (1962). A personal reaction to two programed textbooks, *Theory Into Practice*, 1 (1), 7-8.

Luangrath, P., Pettersson, S., & Benckert, S. (2011). On the use of two versions of the Force Concept Inventory to test conceptual understanding of mechanics in Lao PDR. *Eurasia Journal of Mathematics, Science and Technology Education*, *7*(2), 103-114.

Madrazo, G. M. (1997). Using trade books to teach and learn science. *Science and Children*, *34*, 20–21.

Mangen, A., & Kuiken, D. (2014). Lost in an iPad: Narrative engagement on paper and tablet. Scientific Study of Literature, 4(2), 150-177.

Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. International Journal of Educational Research, 58, 61-68.

Maria, K., & Johnson, J. M. (1989, December). Correcting misconceptions: Effects of type of text. Paper presented at the annual meeting of the National Reading Conference, Austin, TX.

Matthews, M. R. (2015). *Science teaching: The contribution of history and philosophy of science* (20th anniversary revised and expanded edition). Oxford: Routledge.

Mayer, D. A. (1995). How can we best use children's literature in teaching science concepts? *Science and Children*, *32*(6), 16–19.

Mercer, N. (2007). Commentary on the reconciliation of cognitive and sociocultural accounts of conceptual change. *Educational Psychologist*, 42(1), 75–78. https://doi.org/10.1080/00461520709336920

Millar, R. & Osborne, J. (1998), *Beyond 2000: Science Education for the Future.* London: Nuffield Foundation.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional science*, *32*(1-2), 99-113.

Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology*, *92*(4), 724.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology*,*96*(1), 165.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of educational psychology*, *97*(1), 117.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and instruction*, *19*(2), 177-213.

Mulop, N., Yusof, K. M., & Tasir, Z. (2014, April). The improvement of confidence level of students learning thermodynamics through a multimedia courseware. In *2014 IEEE Global Engineering Education Conference (EDUCON)* (pp. 733-738). IEEE.

Newton, H (1728). Two letters from Humphrey Newton to John Conduitt. http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00033] [Accessed on 1st May, 2016]

Newton, I (2013). Newton's Waste Book (Part 1). http://www.newtonproject.sussex.ac.uk/view/texts/normalized/NATP00220] [Accessed on 28th April, 2016]

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, *15*(5), 625-632.

Norris, S. P., Guilbert, S. M., Smith, M. L., Hakimelahi, S., & Phillips, L. M. (2005). A theoretical framework for narrative explanation in science. *Science Education*, *89*(4), 535-563.

Özdemir, G., & Clark, D. B. (2007). An overview of conceptual change theories. *Eurasia Journal of Mathematics, Science & Technology Education*, *3*(4), 351-361.

Özdemir, G., & Clark, D. (2009). Knowledge structure coherence in Turkish students' understanding of force. *Journal of Research in Science Teaching*,*46*(5), 570-596.

Paxton, R. J. (1997). Someone with like a life wrote it: The effects of a visible author on high school history students. *Journal of Educational Psychology*, *89*(2), 235.

Paxton, R. J. (2002). The influence of author visibility on high school students solving a historical problem. *Cognition and Instruction*, *20*(2), 197-248.

Penney, K., Norris, S. P., Phillips, L. M., & Clark, G. (2003). The anatomy of junior high school science textbooks: An analysis of textual characteristics and a comparison to media reports of science. *Canadian Journal of Math, Science & Technology Education*, *3*(4), 415-436.

Rice, D. C. (2002). Using trade books in teaching elementary science: Facts and fallacies. *The Reading Teacher*, *55*, 552–565.

Ryan, R. M., Koestner, R., & Deci, E. L. (1991). Ego-involved persistence: When free-choice behavior is not intrinsically motivated. *Motivation and emotion*, *15*(3), 185-205.

Savinainen, A., & Scott, P. (2002). The Force Concept Inventory: a tool for monitoring student learning. *Physics Education*, *37*(1), 45.

Schraw, G. (1998). Processing and recall differences among selective details. *Journal of Educational Psychology*, 90(1), 3.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review*,*13*(1), 23-52.

Shoemaker, C. A. (2010). Student confidence as a measure of learning in an undergraduate Principles of Horticultural Science course. *HortTechnology*, *20*(4), 683-688.

Southerland, S.A., Abrams, E., Cummins, C.L. & Anzelmo , J. (2001). Understanding students' explanations of biological phenomena: Conceptual frameworks or p-prims. *Science Education*, *85*, 311-327.

Spargo, P. E. (2005). Investigating the site of Newton's laboratory in Trinity College. *South African Journal of Science*, *101*, 315-321.

Steinberg, M. S., Brown, D. E., & Clement, J. (1990). Genius is not immune to persistent misconceptions: conceptual difficulties impeding Isaac Newton and contemporary physics students. *International Journal of Science Education*, *12*(3), 265-273.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research* (Vol. 15). Newbury Park, CA: Sage

Sullivan, G. M., & Artino Jr, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of graduate medical education*, *5*(4), 541-542.

Sullivan, G. M., & Feinn, R. (2012). Using effect size-or why the P value is not enough. *Journal of graduate medical education*, *4*(3), 279-282.

Thaden-Koach, T.C., Dufresne, R.J., & Mestre, J.P. (2006). Coordination of knowledge in judging animated motion, *Physics Education Research*, 2, 020107.

Toolan, M. (2001). *Narrative: A critical linguistic introduction*. London: Routledge.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159-169.

Voss, J. F., Wiley, J., & Sandak, R. (1999). On the use of narrative as argument. *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso*, 235-252.

Wade, S. E., & Adams, R. B. (1990). Effects of importance and interest on recall of biographical text. *Journal of Reading Behavior*, 22, 331–353.

Wagner, J.F. (2006). Transfer in pieces. *Cognition and Instruction*, 21, 1-71.

Willingham, D. T. (2004), Ask the cognitive scientist: the privileged status of story. *American Educator*, 28 (4).

Westfall, R. (1983). *Never at rest: a biography of Isaac Newton*. Cambridge: Cambridge University Press.

Willingham, D. T. (2009), *Why Don't Students Like School?* San Francisco, CA: Jossey-Bass.

Zabrucky, K. M. & Moore, D. (1999). Influence of text genre on adults' monitoring of understanding and recall. *Educational Gerontology*, *25*(8), 691-710.

Zaromb, F. M., & Roediger, H. L. (2009). The effects of "effort after meaning" on recall: Differences in within-and between-subjects designs. *Memory & cognition*, *37*(4), 447-463.