



HAL
open science

Supporting conceptual change in school science: A possible role for tacit understanding

Christine Howe, Amy Devine, Joana Taylor Tavares

► To cite this version:

Christine Howe, Amy Devine, Joana Taylor Tavares. Supporting conceptual change in school science: A possible role for tacit understanding. *International Journal of Science Education*, 2011, pp.1. 10.1080/09500693.2011.585353 . hal-00719063

HAL Id: hal-00719063

<https://hal.science/hal-00719063>

Submitted on 19 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Supporting conceptual change in school science: A possible role for tacit understanding

Journal:	<i>International Journal of Science Education</i>
Manuscript ID:	TSED-2010-0457-A.R1
Manuscript Type:	Research Paper
Keywords :	conceptual change, science education, experimental study
Keywords (user):	conceptual change, science education, object fall

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supporting conceptual change in school science: A possible role for tacit understanding

Abstract

When students reason during school science, they often refer to conceptions that are derived from out-of-school experiences and are poor proxies for science orthodoxy. However, for some areas of science, these conceptions represent only a proportion of students' full conceptual knowledge, for tacit understanding exists that is superior to the understanding displayed when reasoning. Noting that tacit understanding is engaged when events are judged as natural or non-natural, the paper is concerned with software that: a) requires the direction and speed of falling objects to be predicted, i.e. a typical science reasoning task that engages conceptual knowledge; b) presents simulations of predicted motion in the expectation that its naturalness or non-naturalness would be recognised. An evaluation study is reported where children aged 8 to 12 years worked with the software in contexts that typify computer use in classrooms, i.e. individually under adult guidance ($n = 44$ children) or in pairs with a classmate ($n = 48$ children). They were observed while they did this. Reasoning about object fall was assessed via change from individual pre-tests administered prior to software usage to individual post-tests administered a few weeks afterwards. The children who worked with the software showed greater pre- to post-test gains in conceptual understanding than control children ($n = 47$ children), who lacked software experience. The gains were especially marked for the children who worked in pairs. The approach taken is contrasted with traditional approaches to conceptual change in school science, e.g. strategies that rely upon classroom experiments.

Introduction

It is now beyond doubt that when students reason during school science, they often refer to conceptions that are derived from out-of-school experiences and are poor proxies for science orthodoxy. For instance, reference to unorthodox conceptions is documented in a significant proportion of the c.8000 entries that comprise Duit's (2007) bibliography. Scholars vary over the terms that they use to characterise these conceptions: misconceptions, prior conceptions, alternative conceptions, alternative frameworks, everyday theories and intuitive notions are a sample of the numerous phrases that have traditionally been used (see, e.g., Driver & Erickson, 1983). Scholars also differ over whether they regard the conceptions as organised into tightly structured theories (McCloskey, 1983; Vosniadou, 1994), loose conglomerates (diSessa, 1993; Wagner, 2006), or a mixture of both (Howe, 1998; Oliva, 1999). However, reference to unorthodox conceptions during reasoning in science is an established fact, and nowadays is widely recognised as a challenge that educators must address. Science teaching should, in other words, support processes of *change* from the use of unorthodox conceptions during reasoning to the use of received ones. Taking the need for change as established, the primary aim behind this paper is to contribute a new perspective upon how change can be supported.

As Scott, Asoko and Leach (2007) point out, contemporary approaches to conceptual change are heavily influenced by the work of Posner, Strike, Hewson and Gertzog (1982). According to Posner et al., effecting change depends upon fulfilment of four conditions: a) there must be dissatisfaction with existing conceptions; b) new conceptions must be minimally understood; c) new conceptions must appear initially plausible; d) new conceptions should suggest the possibility of a fruitful research programme. The fourth condition is probably of marginal significance in school

1
2
3 contexts, where students seldom regard themselves as on the threshold of research
4
5 programmes. However, it is hard to imagine compliance with the first three conditions
6
7 not being necessary, and so there have been numerous attempts to incorporate them
8
9 into teaching programmes. Typically the programmes require students to conduct
10
11 experiments to test their initial conceptions (Harlen, 1999; Hodson, 1993; Scott et al.,
12
13 2007). It is anticipated that results will, on the one hand, trigger dissatisfaction with
14
15 existing conceptions through their failure to provide support, and on the other hand,
16
17 render the target science both comprehensible and plausible through compatible data.
18
19 In reality, research evaluating the approach has obtained patchy results, with many
20
21 studies reporting minimal impact upon conceptual understanding (Chinn & Brewer,
22
23 1993; Duit, 1999; Duit, Treagust, & Widodo, 2008; Harlen, 1999; Limón, 2001).
24
25
26
27
28
29

30 Discussions of why evaluations have proved disappointing typically revolve
31
32 around three types of explanation: a) experiments often fail to generate the requisite
33
34 data; b) Posner et al.'s three conditions may be necessary but they are insufficient; c)
35
36 cognitive limitations eclipse the significance of experimental evidence. Experimental
37
38 failure is graphically and comprehensively documented in Claxton (1991), with
39
40 research indicating that inadequate control of variables is often responsible (Howe,
41
42 Tolmie, & Sofroniou, 1999; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Schauble,
43
44 1990; Scottish Executive Education Department, 2005). As regards possible
45
46 insufficiencies in Posner et al.'s conditions, social-constructivist critiques (see Duit,
47
48 1999; Duit et al., 2008; Limón, 2001) show how cultural (primarily discursive)
49
50 practices shape the significance accorded to experiments. A specific illustration
51
52 appears in Howe, Rodgers, and Tolmie (1990): when 8- to 12-year-old children
53
54 conducted experiments in small groups, any value from evidence regarding which
55
56 objects floated or sank depended on the discussions that the evidence generated. In
57
58
59
60

1
2
3 particular, when groups debated a range of ideas about how the evidence should be
4
5 interpreted, group members showed improved understanding between individual pre-
6
7 tests prior to group work and individual post-tests a few weeks later. When groups
8
9 held similar ideas and debate was thereby precluded, conceptual understanding was
10
11 identical at pre- and post-test, despite access to equivalent experimental evidence.
12
13
14

15
16 As regards cognitive limitations, difficulties have been identified at all ages
17
18 with adopting the meta-cognitive perspective required to link conceptions with
19
20 experimental data (Driver, Leach, Millar, & Scott, 1996; Kuhn, 2001). Conceptions *or*
21
22 data can be brought into focus but seldom both together. Meta-cognitive difficulties of
23
24 this kind also help to explain the ‘confirmation biases’ that have been reported, i.e. the
25
26 apparent discounting of experimental results that counter preconceptions (e.g. Dunbar
27
28 & Klahr, 1989; Koslowski, 1996; Mynatt, Doherty, & Tweney, 1978). This said,
29
30 discounting also occurs when the conceptual implications of discordant results are
31
32 recognised, suggesting that confirmation biases need to be acknowledged as a
33
34 limitation in their own right. For instance, many of the 9- to 12-year olds reported in
35
36 Howe and Tolmie (2003) expressed surprise at different rates of cooling from black
37
38 and white containers because they ‘knew’ colour makes no difference. They usually
39
40 attributed their results to defective thermometers (see Chinn & Brewer, 1993, for
41
42 equivalent examples with other topics). Moreover, neuroscience research
43
44 demonstrates how data that confirm preconceptions and data that disconfirm both
45
46 result in brain activation (Fugelsang & Dunbar, 2005). The key difference is that only
47
48 confirmatory data activate brain areas associated with learning; disconfirmatory data
49
50 activate areas associated with error detection.
51
52
53
54
55
56
57
58

59
60 Nevertheless, while the problems must be recognised, it is unclear whether
they should be interpreted as challenging Posner et al.’s (1982) conditions or

1
2
3 challenging the use of experiments to fulfil those conditions. The latter is plausible for
4
5 at least two reasons. First, when experiments are used in accordance with Posner et
6
7 al.'s conditions, their results have to be treated *simultaneously* as counter-evidence on
8
9 initial conceptions and indicative evidence for target alternatives. Yet the cognitive
10
11 literature contains numerous illustrations of difficulties with considering single events
12
13 from multiple perspectives (e.g. Flavell, Miller, & Miller, 1993). Karmiloff-Smith
14
15 (1992) reiterates the point in the specific context of science. Second, if experiments
16
17 are sufficiently robust to be consistent with curriculum targets, they will, of necessity,
18
19 be decontextualised from the lifetime of social and physical experiences that support
20
21 initial conceptions. They may therefore appear isolated and meaningless rather than
22
23 momentous. So the question is raised of whether an alternative to experiments can be
24
25 identified to fulfil Posner et al.'s conditions, and if it can be whether it has positive
26
27 implications for conceptual change. The remainder of this paper attempts to
28
29 demonstrate that, for some areas of science at least, the answer may be affirmative.
30
31 Specifically, a non-experimental strategy for promoting conceptual change is
32
33 described that concurs with Posner et al. but: a) separates the challenging of existing
34
35 conceptions from the highlighting of target alternatives; b) calls on experiences that
36
37 are just as familiar and personally meaningful as the ones that support initial
38
39 conceptions. An encouraging evaluation study is then reported.
40
41
42
43
44
45
46
47
48
49
50
51

52 **An alternative strategy for promoting conceptual change**

53 *The potential relevance of tacit understanding*

54
55
56
57
58
59 The strategy stemmed from evidence that, for some areas of science, the conceptions
60
that students call upon when reasoning represent only a proportion of their full

1
2
3 conceptual knowledge. Reasoning, no matter whether it involves planning actions,
4 predicting events or interpreting outcomes, requires *explicit* engagement with
5 conceptual knowledge, i.e. events must be related to underlying conceptions and
6 relations must be reflected on and used to draw inferences. Thus, reasoning tasks are
7 inherently more demanding than tasks, such as much routine action but also
8 judgments of familiarity, that require only *tacit* engagement, i.e. tasks that necessitate
9 relating events to conceptual knowledge, but do not demand reflection or inference.
10 Substantial differences have been demonstrated between the conceptual knowledge
11 displayed when tacit engagement suffices and the conceptual knowledge displayed
12 during reasoning. Unsurprisingly, given the lesser task demands, these differences
13 usually suggest that the former is superior. For instance, Reed, McLeod and Dienes
14 (2010) found that skilled ball players, who reliably make the eye gaze adjustments
15 that allow balls to be caught, show no comprehension of the role of eye gaze when
16 explaining how catching is achieved. Displaying scenarios where rolling cylinders of
17 varying mass collide with a stationary cylinder, Kotovsky and Baillargeon (1998)
18 found that 6-month-old infants show surprise when post-collision speed fails (non-
19 naturally) to be proportionate to the rolling cylinders' mass but not when it is
20 proportionate. Yet Inhelder and Piaget (1958 – later Howe, 1998) found that
21 throughout the primary school age range, children have difficulties reasoning about
22 the distances that objects of varying mass roll along tracks.

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52 It was the ability to differentiate between natural and non-natural events,
53 illustrated in the work of Kotovsky and Baillargeon, which suggested the strategy to
54 be developed below. In principle, the ability seemed to offer opportunities for
55 separating evidence that challenges initial conceptions from evidence that supports
56 curriculum targets *and* for employing evidence that is familiar and meaningful. As
57
58
59
60

1
2
3 regards separation, it seemed possible that, with topics where the ability to
4
5 differentiate exists, events could be displayed that reflect students' erroneous
6
7 reasoning in the expectation that the non-naturalness will be recognised, potentially
8
9 creating dissatisfaction with existing conceptions but not in itself introducing
10
11 alternatives. As an independent yet coherent step, authentic events could be presented
12
13 subsequently, anticipating that their naturalness too will be recognised. Through this
14
15 second step (and in a fashion that mirrors the role intended for experimental results in
16
17 traditional teaching programmes), target conceptions should be highlighted. Because
18
19 the recognition of naturalness and non-naturalness calls upon knowledge which
20
21 students already possess (and may, in some cases, have possessed since infancy), it
22
23 also employs evidence that is fully grounded in past experiences. Such evidence will
24
25 be at least as familiar and personally meaningful as the data that support unorthodox
26
27 conceptions in reasoning. Indeed, it fulfils one of Chinn and Brewer's (1993) main
28
29 criteria for credible evidence, 'data that the individual already believes' (pp.25).
30
31
32
33
34
35

36
37 In short, the suggested strategy involves engaging tacit understanding through
38
39 tasks that require discrimination between natural and non-natural events in the hope of
40
41 boosting the conceptual knowledge that is displayed when reasoning. To provide a
42
43 concrete prototype (and as a prelude to evaluation), the strategy has now been
44
45 embodied in computer software, in particular in a programme that starts by requiring
46
47 events to be predicted, i.e. a reasoning task. It then presents simulations of predicted
48
49 events, followed with simulations of how events proceed in reality and invitations to
50
51 make comparisons. Simulations were essential given the need to present non-natural
52
53 displays when predictions are incorrect, and they have the advantage of sidestepping
54
55 the 'failed experiment' problem outlined above. While questions have been raised
56
57 about whether students treat simulations as genuine representations (Hennessy &
58
59
60

1
2
3 O'Shea, 1993), they have been used successfully in work with 'alternative realities'
4
5 (e.g. Hennessy, Twigger, Driver, O'Shea, O'Malley, Byard, Draper, Hartley,
6
7 Mohamed, & Scanlon, 1995), albeit work that juxtaposes natural and non-natural
8
9 events in complex reasoning tasks rather than uses them merely to stimulate
10
11 recognition of naturalness.
12
13
14

15 16 17 18 19 *Object fall software* 20

21
22 The topic selected for software implementation was patterns of motion when objects
23
24 fall after being dropped from some height. Student reasoning about both the direction
25
26 and speed of motion has been extensively explored, with significant errors reported.
27
28 With direction, the focus has been upon the paths along which objects travel when
29
30 they fall after moving horizontally, e.g. when balls roll over cliffs or litter is dropped
31
32 from moving vehicles (e.g. Anderson, Tolmie, Howe, Mayes, & Mackenzie, 1992;
33
34 Eckstein & Kozhevnikov, 1997; Eckstein & Shemesh, 1989; Hood, 1995; Kaiser,
35
36 Profitt, & McCloskey 1985; Krist, 2000; Marioni, 1989; McCloskey 1983; Whitaker,
37
38 1983). Regardless of student age (from preschool- to undergraduate-level), the main
39
40 finding is that objects are variously predicted to fall vertically, travel backwards, fall
41
42 diagonally forwards, or continue horizontally in space before falling vertically.
43
44 However, they are seldom predicted to trace the parabolic paths in a forward direction
45
46 that they actually follow. With speed, the emphasis has been upon fall from rest, with
47
48 two main messages that again apply across a wide age range (e.g. Baker, Murray, &
49
50 Hood, 2009; Cahyadi & Butler, 2004; Champagne, Klopfer, & Anderson, 1980;
51
52 Chinn & Malhotra, 2002; Gunstone & White, 1981; Hast & Howe, 2009; Howe,
53
54 1998; Nachtigall, 1982; van Hise, 1988). The first is that when objects vary only in
55
56
57
58
59
60

1
2
3 mass heavy items are typically predicted to fall faster than light items rather than
4
5 travel at speeds that, even taking air resistance into account, are usually almost
6
7 identical. Another is that, regardless of mass, objects are expected to reach maximum
8
9 velocity quickly, and then fall with constant or decelerating velocity.
10
11

12
13
14 At the same time, there is evidence that students are capable of differentiating
15
16 between natural and non-natural displays of object fall in precisely the contexts that
17
18 are associated with faulty reasoning. For instance, Kaiser, Profitt, Whelan and Hecht
19
20 (1992) found that when undergraduates viewed computer-simulated kegs falling from
21
22 aeroplanes, they consistently judged forward parabolas as natural and other
23
24 trajectories as non-natural. Kim and Spelke (1999) obtained equivalent results with
25
26 preschool children from two years of age, using judgment tasks and tasks that
27
28 involved monitoring of eye gaze as indices of surprise. When Shanon (1976)
29
30 presented videotapes of balls falling with constant or accelerating velocity to an
31
32 undergraduate sample, he found that constant velocity was consistently judged as non-
33
34 natural, while acceleration was regarded as natural. Moreover, using a series of
35
36 computer-simulations, the present authors [**Reference to be included after review**]
37
38 found primary school children capable of differentiating natural direction and speed
39
40 from non-natural direction and speed during fall: a) from rest; b) after horizontal
41
42 motion; c) through air alone; d) through air and then water. The children were much
43
44 less successful at predicting direction and speed with comparable scenarios.
45
46
47
48
49
50

51
52 With evidence across a wide age range for tacit understanding that surpasses
53
54 the understanding displayed in reasoning, object fall was regarded as an appropriate
55
56 (and educationally relevant) topic. Accordingly, the software, which was programmed
57
58 using Macromedia Director, used scenarios that showed a girl in a hot air balloon,
59
60 holding and then dropping a ball, these scenarios having been employed in the

1
2
3 authors' preceding research. Sixteen scenarios were used, organised into two blocks
4
5 of eight. One block addressed the direction of fall, and the other addressed changes in
6
7 speed¹. Within each block, the scenarios depicted all possible combinations of
8
9 whether: a) the balloon was stationary or moving at the time the ball was released; b)
10
11 the ball fell through air alone or through air and then into a swimming pool; c) a green
12
13 or silver ball was used, with the green ball shown via a real equivalent to be very light
14
15 and the silver ball (of identical size) shown to be extremely heavy. Each time the task
16
17 was presented, the order of blocks and order of scenarios within blocks were
18
19 randomly varied via the computer programme. At the start of each scenario, an
20
21 instruction to notice which ball was being used appeared beside a close-up of the girl
22
23 and ball (see Figure 1a). Clicking 'Ready' caused the picture to zoom out so the
24
25 whole scenario was visible. An instruction to notice if the balloon moves appeared
26
27 (see Figure 1b), with clicking 'Go' activating the scenario, i.e. with stationary
28
29 scenarios the ball was released and with moving scenarios the balloon moved part
30
31 way across the screen prior to release. [Insert Figure 1]
32
33
34
35
36
37
38

39
40 At the moment of release, the action froze, and predictions of subsequent
41
42 motion were invited. With the direction scenarios, the prediction phase began with the
43
44 appearance of three small white circles under the ball: a) directly below; b) behind; c)
45
46 in front (see Figure 1c). This was accompanied with an instruction to select the point
47
48 that the ball would travel through. Once a circle was selected, this turned red and the
49
50 other circles disappeared. At the same time three further white circles appeared below
51
52 the selected circle: a) if the selected circle was directly under the ball, the new circles
53
54 were directly below, behind, and in front; b) if the selected circle was behind the ball,
55
56 the new circles were directly below this circle, parabolically behind and diagonally
57
58 behind; c) if the selected circle was in front of the ball, the new circles were directly
59
60

1
2
3 below, parabolically in front and diagonally in front (see Figure 1d). The instruction
4 to select a point also re-appeared. Once a second circle was chosen this too turned red,
5
6 the other circles disappeared, and three further circles appeared below the second
7
8 circle in the same relative positions, accompanied with the instruction to select. While
9
10 the first and second circles were always located in air, the third set varied with
11
12 medium, i.e. just above the grass or under the water's surface (just below with the
13
14 green ball depicting where it would naturally float; on the pool's floor with the silver
15
16 ball reflecting the fact it would sink).
17
18
19
20
21

22
23 With the speed scenarios, the prediction phase involved answering questions,
24
25 with response options (the words that are capitalised below) presented under each
26
27 question and selected by clicking with the computer mouse. Order of options was
28
29 varied randomly each time the task was presented. With scenarios where the ball fell
30
31 only through air, questioning began with: 'As the ball falls through the air will the
32
33 speed CHANGE or stay the SAME?' If this question was answered correctly via
34
35 CHANGE, a second question followed: 'As the ball falls will its speed get SLOWER
36
37 or FASTER?' There was no second question if the initial question was answered
38
39 incorrectly via SAME. With scenarios where the ball fell through air and water,
40
41 questioning began with: 'Will the ball fall at the SAME speed through the air and the
42
43 water or at DIFFERENT speeds?' The follow-up question after correct selection of
44
45 DIFFERENT was: 'Will the ball travel slower/faster through the AIR or the
46
47 WATER?' with use of 'slower' or 'faster' determined randomly for each presentation.
48
49 Once more, there was no second question after incorrect selection of SAME.
50
51
52
53
54
55

56 With both the direction and speed scenarios, correct predictions triggered the
57
58 message 'Well done! You are correct' inside a gold star to the right of the screen. This
59
60 was accompanied with 'Click upon GO to see the ball fall', with clicking activating a

1
2
3 simulation of the ball falling naturally. Incorrect predictions triggered 'Click on GO to
4 see what you thought would happen', with clicking this time activating a simulation
5 of the ball falling non-naturally in the predicted fashion. Once the latter simulation
6 was complete, 'Did it look correct?' appeared to the right of the screen, with YES and
7 NO as response options. Selection of YES was not expected to occur very often, but if
8 it was chosen 'Sorry the answer is WRONG' appeared on the right of the screen
9 accompanied with 'Click on NEXT to see what really happens'. Selection of NO
10 triggered 'Yes, the motion of the ball was incorrect', together with the invitation to
11 see what really happens. In any event, indicating that the incorrect motion was in fact
12 incorrect triggered 'What is the difference between the incorrect and correct motion?'
13 and then 'Why are they different?' The option was available of replaying the incorrect
14 and/or correct motion as many times as was deemed necessary to answer the
15 questions. The computer programme was designed to record details of button presses
16 at all stages, including predictions.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 Method

41 *Overview*

42
43 With the strategy implemented in software, the primary issue was whether the
44 software had positive implications for conceptual change, and this was the issue that
45 the evaluation study was intended to resolve. It was recognised that a positive
46 outcome would raise further questions, e.g. whether the software was effective
47 *because* it implemented the strategy, but it seemed advisable to address one issue at a
48 time. Accordingly, after securing university ethical approval, a three-stage study was
49 conducted with a sample of 8- to 12-year-old children. This age group was selected
50
51
52
53
54
55
56
57
58
59
60

1
2
3 because: a) the topic area is consistent with curriculum requirements for the 8- to 12-
4 year range (DfEE/QCA, 1999); b) the authors' previous research [**Reference to be**
5
6
7
8 **included**] indicates that the materials are comprehensible for children in this age
9 range, and interest can be maintained across 16 scenarios.
10
11

12
13
14 During the first stage of the study, the sample was pre-tested to establish base-
15 line reasoning about object fall. For the second stage, about two-thirds of the pre-
16 tested sample were assigned to work with the software, with the remaining third
17 intended as an untutored control. Since the evaluation was designed merely to
18 ascertain whether the software supported conceptual growth, comparison with an
19 untutored control group was appropriate, rather than, for instance, comparison with a
20 group that studied object fall in an alternative fashion. Half of the children selected to
21 work with the software (i.e. one third of the pre-tested sample) did this individually,
22 with an adult on hand to highlight relevant features and offer guidance upon request.
23 The other half worked with a classmate, with the adult offering guidance upon request
24 but otherwise being deliberately passive. As noted, social-constructivist analyses have
25 drawn attention to the cultural context in which procedures derived from Posner et al.
26 (1982) are situated, emphasising discourse practices. While exploring cultural
27 influences was not a primary aim, it was recognised that the software could not be
28 presented in a contextually 'neutral' fashion. Accordingly, an attempt was made to
29 mimic two contexts, which not only occur frequently during computer-based sessions
30 in classrooms (Crook, 1994) but are also known to be associated with markedly
31 different patterns of discourse (Anderson, Tolmie, McAteer, & Demissie, 1993). The
32 final stage of the study involved a post-test, which was administered to the children
33 who worked with the software and to the control group.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Sample

The sample was recruited from Years 4 to 6 of a state-sector primary school located in a predominantly lower middle-class area of Cambridge, UK. With every parent consenting to participation, all children in the relevant age range were potentially included. However, absence from school meant that six children completed only one of the pre- and post-tests. With these children excluded, the sample comprised 139 children (77 girls). Forty-eight children were from Year 4, 47 were from Year 5, and 44 were from Year 6, with a sample mean age of 9.73 years ($SD = 0.95$).

Pre- and post-tests

The pre- and post-tests each comprised 16 items, which were identical across the two tests. Each item was associated with one of three topics: a) a ball falling from a hot air balloon (as with the software), $n = 5$ items; b) a box falling from a helicopter, $n = 6$ items; c) a box falling from a train on a high bridge, $n = 5$ items. The items were selected from a pool of 32 items that had previously been piloted with a similar aged (but different) sample of children. Rasch analysis of pilot predictions (Bond & Fox, 2007) was used to identify items of varying difficulty, whose 'fit values' were nevertheless consistent with a uni-dimensional scale. Sufficient items fulfilled these criteria to permit the selection of eight items that addressed the direction of fall and eight items that addressed speed change, while also ensuring that each set of eight covered all possible combinations of fall from rest vs. after motion, fall onto grass vs. into water, and fall of a heavy vs. light object. To facilitate presentation, the items were organised into two blocks, with one block addressing direction and the other block addressing speed. There was a fixed order of blocks during presentation and a

1
2
3 fixed order of combinations within blocks, although both orders were initially
4
5 determined at random. For presentation purposes, each test item comprised a coloured
6
7 illustration of the scenario accompanied with text.
8
9

10
11 With each of the eight direction items, the text included instructions to
12
13 indicate where the object lands by marking the illustration with a cross and then to
14
15 draw the path that the object follows as it falls. With the four speed items that
16
17 addressed fall onto grass, the text was ‘As the ball/box falls through the air, will the
18
19 speed of the ball/box (varied as appropriate for the scenario)?’ with ‘Get faster’, ‘Get
20
21 slower’ and ‘Stay the same’ listed below as possible answers. With the four items that
22
23 addressed fall into water, the text was ‘Will the ball/box fall at the same speed
24
25 through the air and the water or at different speeds?’ again accompanied with three
26
27 possible answers, ‘Faster in the air’, ‘Faster in the water’ and ‘Same speed in air and
28
29 water’. In all eight cases, the possible answers to the speed items were presented in
30
31 randomly varying orders as multiple-choice options, with response boxes to tick.
32
33 Four direction items and four speed items also invited explanations of predicted
34
35 paths/speeds, offering six multiple-choice options (again in randomly varying orders)
36
37 as possible factors: ‘A force pulling the ball/box downwards’, ‘The balloon/helicopter
38
39 /train’s direction before the ball/box started falling’, ‘The air or water pushing the
40
41 ball/box upwards’, ‘There is no wind’ (absence of wind was highlighted in the task
42
43 instructions – see below), ‘The air or water pulling the ball/box downwards’, ‘The
44
45 ball/box’s weight’. With the direction items, the first two of the above factors were
46
47 regarded as relevant, and the remaining four as irrelevant (but indicated in the
48
49 background literature as often used in reasoning). With the speed items, the first and
50
51 third factors were regarded as relevant, and the remaining four as irrelevant.
52
53
54
55
56
57
58
59
60

Procedure

In accordance with the school's preference, the children completed the pre-test in year groups, i.e. two classes together, sometimes in the library and sometimes in the room normally used by one of the classes. Each child was issued with an answer booklet that contained the test items in sequence (i.e. coloured illustrations and text), and invited to insert their name, school, class and gender on the front cover, together with the date. Thereafter, a researcher took the group through the items, displaying them in sequence on a large screen using PowerPoint, ensuring that the children were looking at the corresponding item in their booklets, talking them through what the task involved, and inviting them to enter responses in their booklets. The researcher also highlighted features that were hard to discern from the booklets alone, i.e. whether the carrier was stationary or moving prior to the ball/box's release, whether the ball/box was light or very heavy, and the intended absence of wind. The class teachers were always present during the pre-test to assist with procedural questions and to ensure that there was no copying of neighbours (but not to help with item solution). The test took approximately 25 minutes to complete.

Roughly one third of each class was assigned at random to each of the second-stage conditions, i.e. individually but with adult guidance (designated, for brevity, the 'single' condition), with a classmate but with minimal adult guidance (designated the 'pair' condition), and untutored (designated the 'control' condition). This resulted in 44 children being placed in the single condition, 48 being placed in the pair condition and 47 being placed in the control group, with no significant differences between the conditions over gender composition ($X^2(2) = 1.44, ns$) or school year ($X^2(4) = 0.17, ns$). The children assigned to the pair and single conditions came to a private room at their school, where the researcher who had administered the pre-test welcomed them,

1
2
3 provided an overview of the procedure, invited participation and, with consent to
4
5 continue (given in all cases), obtained basic demographic information. Thereafter she
6
7 directed the children to a Dell Latitude D820 laptop computer, and showed them how
8
9 to respond using the computer mouse.
10
11

12
13 With the children who were working in pairs, the researcher then withdrew to
14
15 another part of the room, only intervening on the rare occasions that she was asked for
16
17 procedural advice. On the other hand, she stayed with the children who were in the
18
19 single condition, again offering procedural assistance upon request but this time also
20
21 highlighting key features. For instance, she asked the children to describe and explain
22
23 the difference between the incorrect and correct motion after these had been viewed,
24
25 scaffolding the use of appropriate explanatory factors. Regardless of condition, the
26
27 researcher used a grid to record whether each software scenario was associated with:
28
29 a) discussion of the prediction; b) discussion of the simulated motion; c) analysis of
30
31 differences between correct and predicted motion; d) interventions from the
32
33 researcher. Brief notes were made of discourse content. Upon completion of the
34
35 computer-presented task (which took between 20 and 30 minutes), all children wrote
36
37 individual answers to two questions: a) when an object falls what makes a difference
38
39 to the path that it travels through as it falls; b) what makes a difference to the speed of
40
41 an object when it falls? Finally, between two and four weeks after the single and pair
42
43 children worked with the software, the post-test was administered, following
44
45 procedures that were identical to the pre-test.
46
47
48
49
50
51
52

53 54 55 56 57 *Scoring*

58
59 Five measures of conceptual knowledge were obtained from the pre- and post-tests:
60

- Correct predictions - cross (CP/C) = Number of direction items where the cross was positioned at the correct landing point (Maximum score possible = 8.00, with eight direction items)
- Correct predictions - path (CP/P) = Number of direction items where the path from the balloon to the cross was drawn correctly (Maximum score possible = 8.00, with eight direction items)
- Correct predictions - speed (CP/S) = Number of speed items where both prediction questions were answered correctly (Maximum score possible = 8.00, with eight speed items)
- Correct explanations (CE) = Number of correct factors selected to explain predictions across the direction and speed items (Maximum score possible = 16.00, with explanations solicited for eight items and two correct options provided with each item)
- Incorrect explanations (IE) = Number of incorrect factors selected to explain predictions across the direction and speed items (Maximum score possible = 32.00, with explanations solicited for eight items and four incorrect options provided with each item).

Results

As noted, the pre- and post-test items were selected from a larger pool, which had been tested with a pilot sample. Rasch analysis of the *predictions* made during piloting was used to create a uni-dimensional scale of varying item difficulty. As a result, the CP/C, CP/P and CP/S measures can be assumed to be reliable. On the other hand, *explanations* were not subject to Rasch analysis during scale development (even

1
2
3 though explanation items were also used during piloting), and therefore the reliability
4
5 of the CE and IE measures could not be assumed a priori. As a result, Cronbach's
6
7 alpha was calculated for the CE and IE measures as a post hoc test of reliability.
8
9

10 Acceptable outcomes were achieved (for CE, $\alpha = 0.70$; for IE, $\alpha = 0.76$).

11
12 Accordingly, analyses were conducted on the total scores that the children obtained
13
14 on each of the five measures, without any need to examine specific items separately.
15
16
17

18 The key question was whether working with the software proved beneficial.
19
20 To address the question, mixed-model two-way ANOVAs were conducted on the
21
22 CP/C, CP/P, CP/S, CE and IE scores obtained at pre- and post-test, with test (pre,
23
24 post) as a within-participants factor and condition (single, pair, control) as a between-
25
26 participants factor. With the prediction measures (CP/C, CP/P, CP/S), significant
27
28 main effects of test were obtained in all three cases (for CP/C, $F(2, 136) = 52.58, p <$
29
30 0.001 ; for CP/P, $F(2, 136) = 34.62, p < 0.001$; for CP/S, $F(2, 136) = 7.40, p < 0.01$).
31
32 With all three measures, performance was significantly higher at post-test than at pre-
33
34 test. The main effects of condition were non-significant, but with all three measures
35
36 there were significant interactions between test and condition (for CP/C, $F(2, 136) =$
37
38 $8.09, p < 0.001$; for CP/P, $F(2, 136) = 5.18, p < 0.01$; for CP/S, $F(2, 136) = 5.16, p <$
39
40 0.01). One-way ANOVAs were conducted to clarify these interactions. As can be
41
42 seen in Table 1, there were no significant differences between the conditions at pre-
43
44 test on any of the three prediction measures. However, by post-test, all differences
45
46 were statistically significant, with the control group invariably performing below the
47
48 children who worked with the software. In fact, the mean post-test scores obtained by
49
50 the control group were little different from their mean pre-test scores, with mean CP/S
51
52 scores actually declining slightly across the two tests. In other words, it was only the
53
54 children who worked with the software who progressed from pre- to post-test. This
55
56
57
58
59
60

1
2
3 said, while progress was detected in both the single condition and the pair, it was
4 particularly marked with the pairs. **[Insert Table 1]**
5
6
7

8
9 As regards the factors that influence direction and speed, conceptual growth
10 would be signaled by pre- to post-test increases in CE scores and pre- to post-test
11 decreases in IE scores. With this in mind, Table 1 suggests modest benefits from the
12 software for CE scores in the pair condition and for IE scores in the single condition.
13
14 The control group, by contrast, obtained lower CE scores at post-test than pre-test and
15 higher IE scores. Nevertheless, all changes are modest, and the two-way ANOVAs
16 produced no significant main effects of test or condition and no significant
17 interactions between test and condition. Follow-up analyses are not therefore
18 appropriate. Overall then, the software had unmistakable benefits for the conceptual
19 dimensions that were its actual focus (patterns of direction and speed), particularly
20 when the children worked in pairs with a classmate. However, its implications for the
21 underlying determinants of motion were, at best, embryonic.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 Thinking about why it was the CP/C, CP/P and CP/S measures that were
38 associated with the strongest gains, it may be significant that, as Table 1 makes clear,
39 pre-test scores on all three measures were considerably below the maximum possible.
40 Thus, making appropriate predictions was a significant challenge, and this may have
41 focused the children's attention on the prediction element, i.e. inputting the
42 anticipated motion and receiving feedback from the simulations. They may have paid
43 less attention to the stages at which explanations were highlighted, i.e. interpreting
44 why the predicted and correct motion differed and answering the two written
45 questions at the end. Indeed, if the children's focus can be assumed to have been upon
46 predictions, the differences between the single and pair conditions may also become
47 interpretable. Table 2 shows the mean number of scenarios on which the researcher
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 recorded the four discourse categories, i.e. total number of scenarios across each
4
5 condition where the category was recorded divided by 44 (single condition) or 24
6
7 (pair condition). The most striking difference relates to discussion of the predictions,
8
9 which was never recorded in the single condition but was the most frequently used
10
11 discourse category in the pair condition. With 63% of the scenarios around which
12
13 prediction discussion occurred, the focus was on predictions alone, i.e. asserting
14
15 possibilities and confirming or challenging these. With the remaining 37%, there was
16
17 also an attempt to justify positions. In either case, such discussion would have been
18
19 highly salient, if formulating predictions was the children's major focus. Therefore,
20
21 the fact that it occurred only with the pairs could be significant. **[Insert Table 2]**
22
23
24
25
26
27
28
29

30 **Discussion and implications**

31
32
33 The design of the object fall software was informed by three considerations. The first
34
35 was that Posner et al.'s (1982) conditions for effecting conceptual change have
36
37 normally been interpreted as requiring experimental appraisal of initial conceptions. It
38
39 has been assumed that results will simultaneously challenge these conceptions and
40
41 render the target science as comprehensible and plausible. The second consideration
42
43 related to some potentially misleading implications of using experimental appraisal,
44
45 and therefore of its patchy success in supporting conceptual growth. Specifically,
46
47 questions were raised about an approach that necessitates viewing evidence
48
49 simultaneously from two perspectives, and uses data that is inevitably dissociated
50
51 from personal experience. The third consideration was the possibility of engaging
52
53 students' tacit understanding both to separate the creation of dissatisfaction from the
54
55 introduction of alternatives, and to utilise experiences that are familiar and personally
56
57 meaningful. However, while the three considerations informed the software's design,
58
59
60

1
2
3 at no point were they transformed into hypotheses that the evaluation study aspired to
4 test. Rather, the evaluation study was concerned with the effectiveness of materials
5 that embody the third consideration, and therefore relate only indirectly to the first
6 and second. With this recognised, the following discusses what can be inferred from
7 the results regarding the software's effectiveness, then considers the potential
8 relevance of tacit understanding for science education in general, and finally makes
9 brief comments about the first and second considerations.

10
11
12
13
14
15
16
17
18
19
20 As regards the software's effectiveness, the children who worked with it made
21 significantly greater progress than control children in the conceptual understanding
22 underpinning predictions. This was true for both direction of fall and changes in
23 speed. Less progress was however detected for explanatory factors. With reference to
24 pre-test scores (i.e. as shown in Table 1), it was suggested that even the formulation of
25 predictions would have proved challenging for the participating children, and
26 therefore this may have pre-occupied them to the exclusion of background factors.
27 Certainly, if the children were pre-occupied with predictions, it was more likely due
28 to the formidable conceptual content than, say, to practical difficulties with operating
29 the software. As Table 2 shows, 'researcher intervention' occurred relatively
30 infrequently, implying that the children quickly mastered what they were expected to
31 do and found the software easy to use. This of course augurs well for eventual
32 application in classrooms. However, even if conceptual challenge did result in a focus
33 upon predictions, it is unclear whether this is *sufficient* to account for the limited
34 progress with explanatory factors. To explore further, it would be desirable to work
35 with samples whose pre-test scores indicate stronger initial knowledge, perhaps
36 because they had benefited from using the software on previous occasions. Indeed, the
37 implications of repeated software usage would be of general interest.

1
2
3 While the software was generally effective, its benefits were most marked for
4 the children who worked in pairs with a classmate. Progress was detected amongst the
5 children who worked individually under researcher guidance, but to a lesser degree.
6
7 Assuming pre-occupation with predictions, it was proposed earlier that the pairs'
8 superiority might be related to the fact that *discussion* of predictions only occurred
9 between classmates. If discussion was significant, it could have been for at least two
10 reasons. On the one hand, the significance could have stemmed from the exchanges
11 (roughly one-third of the total prediction discussions) where the children offered
12 justifications for their predictions. Justificatory dialogue has been shown on numerous
13 occasions to be beneficial for conceptual understanding (see Howe, 2010, for a review
14 of many studies, including ones relevant for science). On the other hand, the
15 dialogue's content may not have been critical in this particular context. Its very
16 occurrence might have been sufficient to slow the children down and make them think
17 carefully before making predictions. Actually, this second interpretation is probably
18 the most plausible: the justifications usually referred to explanatory factors and
19 therefore had they been crucial for boosting predictions, growth in explanatory
20 understanding might also have been expected (when, as noted, this was negligible). In
21 any event, assuming discussion of predictions was significant for *some* reason, it
22 becomes important to reflect on its non-occurrence in the single condition. It was
23 certainly not precluded in principle, given the researcher's availability for guidance.
24 Indeed, the researcher was approached at other points in the task (see Table 2).
25 Perhaps children regard adult help during initial problem solving as illegitimate.
26
27

28
29 The suggested interplay between task difficulty, discourse patterns and
30 expectations about adult involvement imply a need for caution before extrapolating
31 beyond the object fall context. It remains to be seen whether equivalent software
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 would prove beneficial with topics that students find easier or harder, and whether at
4
5 any level of difficulty (or with any student sample) working with classmates is more
6
7 helpful than working individually with adult support. All such issues are worth
8
9 exploring, for there are likely to be many areas relevant to science where students
10
11 display tacit understanding that surpasses the conceptual frameworks they employ
12
13 when reasoning. Earlier, examples were given relating to horizontal motion as well as
14
15 vertical, and although currently un-researched, it seems likely that tacit understanding
16
17 exists in domains beyond force and motion. Heat transfer and light propagation are
18
19 obvious candidates. Descriptive studies of students' conceptions typically overlook
20
21 the tacit dimension, sometimes even implying that the constructs displayed when
22
23 reasoning *define* conceptual knowledge. Students are said to hold misconceptions,
24
25 when a more accurate claim would be that misconceptions are used with particular
26
27 reasoning tasks. Thus charting tacit understanding requires theoretical re-positioning
28
29 as well as additional research. On the other hand, once such understanding is
30
31 identified, it would be straightforward to design software that emulates the object fall
32
33 structure, and via this obtain a comprehensive picture of how task difficulty, discourse
34
35 patterns and adult involvement interact to determine learning outcomes.
36
37
38
39
40
41
42

43
44 Positive results using such software would confirm and extend the findings
45
46 from the present study, that reasoning tasks, which engage tacit understanding, can
47
48 boost the conceptual level of reasoning itself. However, based on comparisons with
49
50 untutored control groups, results of this kind would not show whether the boosts were
51
52 actual consequences of tacit engagement, just as the present study does not show this
53
54 either. Although the issue was, as noted, regarded as premature for the study, there is
55
56 no reason why it should not be addressed in the future. The best strategy might be to
57
58 compare across a series of partial versions of the present software, e.g. software that
59
60

1
2
3 was equivalent except for requiring recognition of the incorrectness of non-natural
4 events, and equivalent except for requiring recognition of the correctness of natural
5 events. Indeed, such 'decomposition' would have to be taken a step further if the
6 interest was not simply in whether tacit understanding contributes but also in how it
7 contributes. As noted, the engagement of tacit understanding has the potential to
8 separate the creation of dissatisfaction from the introduction of alternatives, *and* to
9 utilise experiences that are familiar and meaningful. Researchers might wish to assess
10 their relative contribution to positive outcomes. While such issues can be flagged here
11 only as matters for future research, there is significance already in the fact that tacit
12 understanding has this twin potential, and that software which accesses this
13 understanding can support conceptual growth. At minimum, it shows that viable
14 alternatives to traditional experimental appraisal do exist, and it is therefore premature
15 to draw pessimistic inferences about conceptual change from studies that examine
16 experimental appraisal to the exclusion of these alternatives. In other words, the study
17 highlights the fact that conclusions to date have been grounded in research that adopts
18 an overly restricted framework, which can and should be addressed.

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42 While the object fall software avoids experimental appraisal as traditionally
43 understood, it remains consistent with Posner et al.'s (1982) conditions. It did, after
44 all, attempt to unsettle existing conceptions and to highlight the comprehensibility and
45 plausibility of target alternatives. Its challenge is not therefore to the broad paradigm
46 that Posner et al. stimulated, but to the equation of this paradigm with experimental
47 appraisal and the conclusions about efficacy that equation has occasioned. Indeed, the
48 software is not merely consistent with teaching within Posner et al.'s framework; it
49 could perhaps also be employed to support such teaching in classrooms. While
50 caution must be exercised before extrapolating from a one-off, researcher-delivered,
51
52
53
54
55
56
57
58
59
60

1
2
3 out-of-class investigation, it is encouraging as regards routine practice that the
4
5 software was easy to use as well as effective. Furthermore, comparison of individual
6
7 usage under adult guidance and usage with a classmate means that a major source of
8
9 discourse variation during authentic computer-based teaching has been addressed. At
10
11 the very least, it would be interesting to examine the consequences of using the
12
13 software and/or equivalent software with other topics within normal teaching. The
14
15 range of potential topics is not unlimited; it would be surprising, for instance, if
16
17 children have much tacit understanding of the earth in space or Mendelian
18
19 inheritance. However, as noted already, a reasonable range is probable, meaning that
20
21 in the fullness of time software that engages tacit understanding might have a
22
23 significant role to play. For the present, all that can be said is that conceptual change
24
25 needs to be supported within school science, experimental appraisal of existing
26
27 conceptions is not an unqualified success, and tacit understanding has the potential to
28
29 take things forward.
30
31
32
33
34
35
36
37
38
39

40 Footnote

41
42
43 ¹ The software can be downloaded for viewing from [website to be included after
44
45 review]. Please click on '2B1' for the direction of fall block, and '2B2' for the speed
46
47 change block.
48
49
50

51 Acknowledgements

52
53
54 The authors wish to thank the Economic and Social Research Council of Great Britain
55
56 for funding the research, Pepi Sarvary for constructing the software, and the
57
58 participating children, their teachers and head teachers for enthusiastic engagement.
59
60

References

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Anderson, A., Tolmie, A., Howe, C.J., Mayes, J.T., & Mackenzie, M. (1992). Mental models of motion. In Y. Rogers, P. Bibby, & A. Rutherford (eds.), *Models in the mind: Theory, perspectives and application* (pp. 59-71). London: Academic Press.
- Anderson, A., Tolmie, A., McAteer, E., & Demissie, A. (1993). Software style and interaction around the microcomputer. *Computers and Education*, 20, 235-250.
- Baker, S.T., Murray, K., & Hood, B.M. (2009). Children's expectations about weight and speed in falling objects: The younger the judge the better? Poster presented at Biennial Meeting of Society for Research in Child Development, Denver, USA.
- Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Cahyadi, M.V., & Butler, P.H. (2004). Undergraduate students' understanding of falling bodies in idealized and real-world situations. *Journal of Research in Science Teaching*, 41, 569-583.
- Champagne, A.B., Klopfer, L.E., & Anderson, J.H. (1980). Factors influencing the learning of classical mechanics. *American Journal of Physics*, 48, 1074-1079.
- Chinn, C.A., & Brewer, W.F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63, 1-49.
- Chinn, C.A., & Malhotra, B.A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, 94, 327-343.
- Claxton, G. (1991). *Educating the inquiring mind: The challenge for school science*. New York: Harvester Wheatsheaf.
- Crook, C. (1994). *Computers and the collaborative experience of learning*. London: Routledge.

- 1
2
3 DfEE/QCA (1999). *Science: The national curriculum for England*. London: HMSO.
- 4
5
6 diSessa, A.A. (1993). Toward an epistemology of physics. *Cognition and Instruction*,
7
8 *10*, 105-225.
- 9
10 Driver, R., & Erickson, G. (1983). Theories-in-action: Some theoretical and empirical
11
12 issues in the study of students' conceptual frameworks in science. *Studies in*
13
14 *Science Education*, *10*, 37-60.
- 15
16
17 Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*.
18
19 Milton Keynes: Open University Press.
- 20
21
22 Duit, R. (1999). Conceptual change approaches to science education. In W. Schnotz,
23
24 S. Vosniadou, & M. Carretero (eds.), *New perspectives on conceptual change*
25
26 (pp. 263-314). Amsterdam: Elsevier.
- 27
28
29 Duit, R. (2007). Bibliography STCSE (Students' and teachers' conceptions and
30
31 science education). <http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html>.
- 32
33
34 Duit, R., Treagust, D.F., & Widodo, A. (2008). Teaching science for conceptual
35
36 change: Theory and practice. In S. Vosniadou (ed.), *International handbook of*
37
38 *research on conceptual change* (pp. 629-646). New York: Routledge.
- 39
40
41 Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery
42
43 processes. In G.M. Green (ed.), *Complex information processing* (pp. 107-143).
44
45 Hillsdale, NJ: Lawrence Erlbaum.
- 46
47
48 Eckstein, S.G., & Kozhevnikov, M. (1997). Parallelism in the development of children's ideas
49
50 and the historical development of projectile motion theories. *International Journal of*
51
52 *Science Education*, *19*, 1057-1073.
- 53
54
55 Eckstein, S.G., & Shemesh, M. (1989). Development of children's ideas on motion:
56
57 Intuition vs. logical thinking. *International Journal of Science Education*, *25*,
58
59 671-688.
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Flavell, J.H., Miller, P.H., & Miller, S.A. (1993). *Cognitive development (Third edition)*. Englewood Cliffs, NJ: Prentice Hall.
- Fugelsang, J.A., & Dunbar, K.N. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia*, *43*, 1204-1213.
- Gunstone, R.F., & White, R.T. (1981). Understanding gravity. *Science Education*, *65*, 291-299.
- Harlen, W. (1999). *Effective teaching of science: A review of research*. Edinburgh: Scottish Council for Research in Education.
- Hast, M., & Howe, C. (2009). Young children's explicit and tacit understanding of object speed and acceleration. Poster presented at BPS Developmental Section Annual Conference, Nottingham, UK.
- Hennessy, S., & O'Shea, T. (1993). Learner perceptions of realism and 'magic' in computer simulation. *British Journal of Educational Technology*, *24*, 125-138.
- Hennessy, S., Twigger, D., Driver, R., O'Shea, T., O'Malley, C. E., Byard, M., Draper S., Hartley, R., Mohamed, R., & Scanlon, E. (1995). A classroom intervention using a computer-augmented curriculum for mechanics. *International Journal of Science Education*, *17*, 189-206.
- Hodson, D. (1993). Re-thinking old ways: Towards a more critical approach to practical work in school science. *Studies in Science Education*, *22*, 85-142.
- Hood, B. M. (1995). Gravity rules for 2- to 4-year-olds? *Cognitive Development*, *10*, 577-598.
- Howe, C. (1998). *Conceptual structure in childhood and adolescence: The case of everyday physics*. London: Routledge.
- Howe, C. (2010). *Peer groups and children's development*. Oxford: Blackwell.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Howe, C., Rodgers, C., & Tolmie, A. (1990). Physics in the primary school: Peer interaction and the understanding of floating and sinking. *European Journal of Psychology of Education, V*, 459-475.
- Howe, C., & Tolmie, A. (2003). Group work in primary school science: Discussion, consensus and guidance from experts. *International Journal of Educational Research, 39*, 51-72.
- Howe, C., Tolmie, A., & Sofroniou, N. (1999). Experimental appraisal of personal beliefs in science: Constraints on performance in the 9 to 14 age group. *British Journal of Educational Psychology, 69*, 243-274.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking*. New York: Basic Books.
- Kaiser, M. K., Profitt, D.R., & McCloskey, M. (1985). The development of beliefs about falling objects. *Perception & Psychophysics, 38*, 533-539.
- Kaiser, M. K., Profitt, D.R., Whelan, S.M., & Hecht, H. (1992). Influences of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 669-690.
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: Bradford Books.
- Kim, I-K., & Spelke, E.S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science, 2*, 339-362.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Kotovskiy, L., & Baillargeon, R. (1998). The development of calibration-based reasoning about collision events in young infants. *Cognition, 67*, 311-351.
- Krist, H. (2000). Development of naïve beliefs about moving objects: The straight-down belief in action. *Cognitive Development, 15*, 281-308.

- 1
2
3 Kuhn, D. (2001). What is scientific thinking and how does it develop? In U. Goswami
4 (ed.), *Blackwell handbook of childhood cognitive development* (pp. 371-393).
5
6 Oxford: Blackwell.
7
8
9
10 Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies in
11 knowledge acquisition. *Monographs of the Society for Research in Child*
12 *Development*, 60, 245.
13
14
15
16
17 Limón, M. (2001). On the cognitive conflict as an instructional strategy for
18 conceptual change: A critical appraisal. *Learning and Instruction*, 11, 357-380.
19
20
21
22 Marioni, C. (1989). Aspects of students' understanding in classroom settings (age 10-17): Case
23 study on motion and inertia. *Physics Education*, 24, 273-277.
24
25
26
27 McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A.L. Stevens
28 (eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Lawrence Erlbaum.
29
30
31 Mynatt, C.R., Doherty, M.E., & Tweney, R.D. (1978). Consequences of confirmation
32 and disconfirmation in a simulated research environment. *Quarterly Journal of*
33 *Experimental Psychology*, 30, 395-406.
34
35
36
37
38
39 Nachtigall, D. (1982). Vorstellungen von Fünftkläßlern über den freien Fall.
40 *Naturwissenschaften im Unterricht - Physik/Chemie*, 30, 91-97.
41
42
43
44 Oliva, J.M. (1999). Structural patterns in students' conceptions of mechanics.
45 *International Journal of Science Education*, 21, 903-920.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Schauble, L. (1990). Belief revision in children. *Journal of Experimental Child Psychology*, 49, 31-57.
- Scott, P., Asoko, H., & Leach, J. (2007). Student conceptions and conceptual learning in science. In S.K. Abell & N.G. Ledermann (eds.), *Handbook of research on science education* (pp.31-56). Mahwah, NJ: Lawrence Erlbaum.
- Scottish Executive Education Department. (2005). *Assessment of achievement programme: Sixth survey of science (2003)*. Edinburgh: Scottish Executive.
- Shanon, B. (1976). Aristotelianism, Newtonianism, and the physics of the layman. *Perception*, 5, 241-243.
- van Hise, Y.A. (1988). Student misconceptions in mechanics: An international problem? *Physics Teacher*, 26, 498-502.
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4, 45-69.
- Wagner, J.F. (2006). Transfer in pieces. *Cognition and Instruction*, 24, 1-71.
- Whitaker, R.J. (1983). Aristotle is not dead: Student understanding of trajectory motion. *American Journal of Physics*, 51, 352-358.

Table 1. Mean Scores at Pre-test and Post-test (*SD* in brackets)

	Pre-test				Post-test			
	Single	Pair	Control	<i>F</i> (2, 136)	Single	Pair	Control	<i>F</i> (2, 136)
CP/C	3.98 (1.27)	4.25 (1.30)	4.42 (1.68)	1.13, <i>ns</i>	5.61 _{ab}	6.04 _b	4.70 _a	5.45, <i>p</i> < .01
CP/P	3.30 (1.21)	3.35 (1.34)	3.49 (1.04)	0.31, <i>ns</i>	4.52 _b	4.64 _b	3.70 _a	4.03, <i>p</i> < .05
CP/S	5.55(1.90)	5.42 (1.92)	5.47 (1.95)	0.05, <i>ns</i>	6.14 _{ab}	6.54 _b	5.21 _a	5.91, <i>p</i> < .01
CE	6.61 (3.21)	6.79 (2.67)	7.23 (2.94)	N/A	6.39	7.31	7.11	N/A
IE	14.57 (5.27)	14.35 (5.03)	14.11 (4.15)	N/A	14.20	14.38	14.57	N/A

Note

When post-test values for CP/C, CP/P and CP/S have different subscripts, values are significantly different (Bonferroni, *p* < 0.01)

Table 2. Mean Number of Scenarios Associated with Each Discourse Category (Maximum possible = 16)

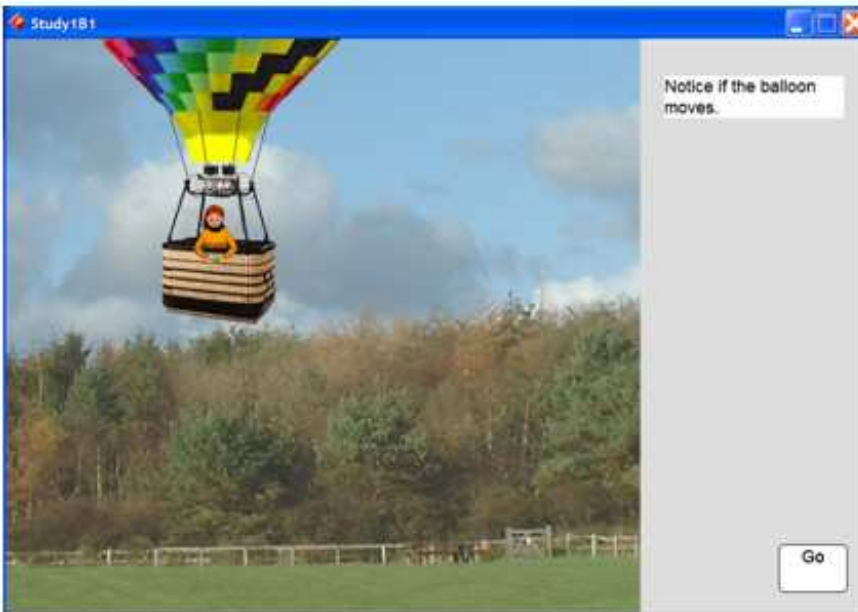
	Single (<i>N</i> = 44 children)	Pair (<i>N</i> = 24 pairs)
Discuss prediction	0	3.00
Discuss simulation	0.25	0.33
Compare correct and incorrect	0.16	0.33
Researcher intervention	1.20	1.38

Figure 1. Sample screen displays

a

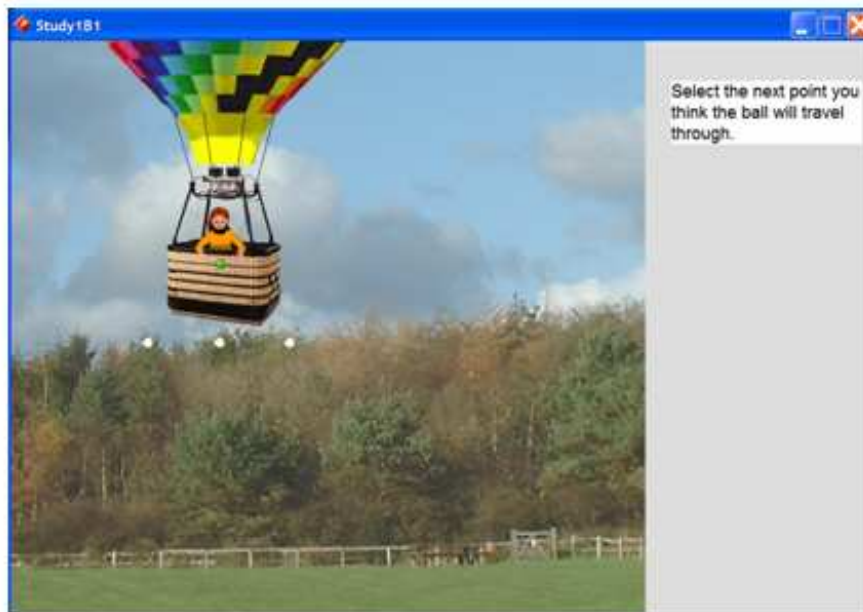


b.



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

c.



d.

