

**SMALL-GROUP COLLABORATION AND INDIVIDUAL KNOWLEDGE
ACQUISITION: THE PROCESSES OF GROWTH DURING ADOLESCENCE
AND EARLY ADULTHOOD**

Christine Howe¹ and Antonia Zachariou²

¹University of Cambridge

²University of Roehampton

Note

The research reported in this manuscript was supported by the British Academy. Thanks are due to the Academy, to Bryony Horsley Heather and Georgina Wong for assistance with data preparation, and to the participating students and their teachers. Correspondence concerning this article should be addressed to Professor Christine Howe, Faculty of Education, University of Cambridge, 184 Hills Road, Cambridge, CB2 8PQ, UK. Email: cjh82@cam.ac.uk.

Abstract

Research into small-group collaboration during middle to late childhood shows that while individual understanding can be promoted through exchanging differing opinions, the joint analyses that groups construct while collaborating play a tangential role. Individuals may or may not accept these constructions depending upon processes of reflection and reconciliation that are triggered through difference and sometimes occur post-group. Recognizing a dearth of research with older participants (together with inconclusive suggestions that collaborative constructions may become more significant with age), the reported study examines the impact of small-group collaboration during adolescence and early adulthood. Forty-six pairs of students aged between 10 and 22 years worked on a computer-presented task that required them to discuss and predict the trajectories objects follow when they fall from stationary or moving carriers. Associations between group dialogue and post-test performance confirmed a key role for differing opinions while collaborative constructions turned out to have little relevance.

Keywords: Small-group collaboration; Knowledge acquisition; Collaborative dialogue; Adolescence; Understanding of object fall

1. Introduction

Research over four decades has provided substantial evidence that individual knowledge and understanding can be promoted through small-group collaboration around problem-solving tasks. The initial evidence (summarized in Doise & Mugny, 1984) related to Piaget's classic 'logico-mathematical problems', especially his conservation and perspective-taking tasks. Slightly later, supportive results were obtained with such social problems as solving moral dilemmas (e.g. Berkowitz, Gibbs, & Broughton, 1980; Leman & Duveen, 1999), distributing rewards (e.g. Damon & Killen, 1982; Kruger, 1992), and dealing with legal transgressions (e.g. Roy & Howe, 1990). Building on both traditions but emphasizing educational relevance, research then began to address curricular tasks, with positive results obtained for music (e.g. Littleton & Mercer, 2012; Miell & Littleton, 2004; Miell & MacDonald, 2000), mathematics (e.g. Damon & Phelps, 1988; Schwarz, Neuman, & Biezuner, 2000), and science (e.g. Azmitia & Montgomery, 1993; Howe, 2010). Contemporaneously, relevance for curriculum mastery was also being demonstrated through numerous studies in the 'cooperative learning' tradition (e.g. Roseth, Johnson, & Johnson, 2008; Slavin & Lake, 2008). Finally and relatively recently, supportive findings have emerged in relation to general thinking tasks, including both spatial matrices (e.g. Mercer & Littleton, 2007) and reasoned argument (Anderson, Howe, Soden, Halliday, & Low, 2001; Fung & Howe, 2014; Vogel, Kollar, Ufer, Reichersdorfer, Reiss, & Fischer, 2016). In all of this work, performance on individual pre-tests prior to group work and individual post-tests upon its completion reveal positive effects from collaborative activity.

At the same time, it has become apparent that individual benefits from group collaboration are context dependent rather than guaranteed. Moreover, of the

contextual factors that have been explored, the one that has been most consistently associated with progress is the exchange through group dialogue of differing opinions about the problems being addressed. Exchanges of this kind have been variously construed as ‘dialectic(al) argumentation’ (e.g. Asterhan & Schwarz, 2009; Larrain, Freire, & Howe, 2014), ‘exploratory talk’ (e.g. Barnes & Todd, 1977; Mercer & Littleton, 2007), ‘interactive engagement’ (e.g. Chi & Wylie, 2014; Vogel et al., 2016), ‘socio-cognitive conflict’ (e.g. Doise & Mugny, 1984; Roy & Howe, 1990), and ‘transactive dialogue’ (e.g. Berkowitz et al., 1980; Miell & MacDonald, 2000). Their relevance is suggested through comparison of groups whose members hold differing opinions with groups whose members hold similar opinions (e.g. Doise & Mugny, 1979; Howe, Rodgers, & Tolmie, 1990; Howe, Tolmie, & Rodgers, 1992b; Mugny & Doise, 1978), for pre- to post-test progress is invariably greatest after the former groups. The suggestion is confirmed through analyses of group dialogue conducted within most of the studies cited so far, for these consistently show pre- to post-test growth to be associated with exchanges around differing opinions.

This pinpointing of difference is of considerable theoretical and practical significance. On the theoretical side, it confirms the emphasis placed in Piaget (1959) upon ‘the pressure of argument and opposition’ (p.137), the explicit endorsement of Piaget’s focus upon ‘a real argument, a real discussion’ in Vygotsky (1998, p.168), and ‘the importance of struggling with another’s discourse’ highlighted in Bakhtin (1981, p.348). From a practical perspective, there is, as detailed in Howe and Mercer (2007), a clear message for teachers as regards: a) the composition of small groups in classrooms (students with different perspectives upon the topic under consideration); b) the design of tasks for group activity (problems and/or problem instructions that draw differences out); c) the nature of preparatory guidance (emphasis on active participation and being unafraid to disagree). Nevertheless, so long as the focus

remains exclusively upon differences of opinion, the theoretical and practical implications will be under-specified, for while differences may be supportive or even necessary they cannot be sufficient. Individual progress must depend also upon productive conclusions being drawn from difference, meaning that for comprehensive analysis the processes underpinning such conclusions need to be understood. The present paper's broadest aim is to contribute to such understanding.

1.1. Collaborative constructions during middle childhood

When the focus (as here) is upon small-group collaboration around problem-solving, some conclusions must typically be drawn during collaboration itself. Unless groups abandon tasks as beyond them, they must construct and converge upon mutually acceptable solutions, i.e. they must produce collaborative constructions. Thus, a key issue is the contribution such constructions make to the individual conclusions on which progress depends. One possibility is that they are the source of these conclusions. In other words, individual conclusions are in effect appropriated group achievements, implying that their productiveness (and therefore the extent of individual progress) is determined by the quality of these achievements. There can be little doubt that this is the process to which Vygotsky subscribed, as evidenced not simply in the famous claim that 'every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level' (Vygotsky, 1978, p.57), but also in the comment on collaborating groups per se that 'the group form of behavior becomes the internal form of behavior' (Vygotsky, 1998, p.169). However, it is possible to envisage alternative processes where individuals play more active roles, reflecting independently upon the differing opinions together perhaps with any collaborative constructions and transforming their understanding in response to these reflections. Piaget presumably thought in these terms given his depiction of

cognitive growth (e.g. Piaget, 1985) as the ‘equilibration’ of conflict between pre-existing knowledge and external experiences. Since Bakhtin’s emphasis was upon how readers derive meaning from ‘discursive struggles’ as presented in novels, he too must have envisaged essentially reflective processes.

As regards progress from group collaboration during middle and late childhood, there is already extensive evidence for the mediating role of individual reflection and reconciliation. To appreciate why, it is important to note that such processes do not preclude close associations between the quality of collaborative constructions and the eventual understanding of individuals (e.g. as reported in Mollard, 2009; Silverman & Geiringer, 1973; Williams & Tolme, 2000), nor do they preclude the appropriation of these constructions en route to growth. It is always possible that upon reflection individuals confirm collaborative constructions as correct and decide to adopt them. Rather, it is *only* reflective processes, and not processes that locate growth within collaborative constructions, that can explain progress in understanding that is not associated with collaborative construction. Yet such progress has frequently been reported in middle and late childhood.

For instance, in much of their research around science, Howe and colleagues found no relation whatsoever between the quality of the ideas that groups of 8- to 12-year-olds converged upon and individual growth from pre- to post-test (e.g. Howe et al., 1990, 1992b; Howe & Tolmie, 2003; Howe, Tolmie, Duchak-Tanner, & Rattray, 2000). Indeed, when the focus of the research was explanatory factors rather than task solutions (e.g. the role of surface friction or slope angle in determining speed down slopes rather than speed itself), groups that actually failed to agree were found to be as effective as groups that achieved consensus (Howe, 2009; Howe et al., 1990).

Likewise, several of the early studies using Piaget’s logico-mathematical problems

indicate explanatory factors being introduced productively at post-test with age groups similar to Howe and colleagues', when these factors had not even been mentioned during group collaboration let alone agreed (see, e.g., Doise & Mackie, 1981). Indeed, with both science and logico-mathematical understanding, there is evidence for progress taking place during the post-group period, albeit stimulated through group dialogue around differing opinions (e.g. Howe, McWilliam, & Cross, 2005; Howe et al., 1992b; Mugny & Doise, 1978; Tolmie, Howe, Mackenzie, & Greer 1993). All in all then, there can be little doubt that when, during middle and late childhood, differences of opinion during small-group collaboration result in growth, individual cognition triggered through difference is heavily involved.

1.2. Adolescence and beyond

The picture is however much hazier with older groups. For sure, there are studies indicating that discussion of differing opinions is as beneficial during adolescence and early adulthood as it is with children. Amongst the reports cited already, Anderson et al. (2001), Asterhan and Schwarz (2009), Berkowitz et al. (1980) and Vogel et al. (2016) all address college or university students, and they all endorse the relevance of difference. Equivalent results are presented in Howe, Tolmie, Anderson, and Mackenzie (1992a), relating to a study where pairs of university undergraduates made collaborative judgments about the speeds of computer-simulated trains: expressed differences were strongly and positively associated with individual progress between pre- and post-tests. Tolmie and Howe (1993) found much the same in research where 12- to 15-year-olds plotted the trajectories that falling objects follow from moving and stationary carriers. Interestingly though, they also detected reluctance amongst mixed-sex groups to debate differences, presumably (especially when this reluctance increased across the age range) reflecting sensitivities amongst teenagers about social

relations. More recently, Jurkowski and Hänze (2015) have shown how trainee teachers' use of 'transactive communication' (which will have included opinion exchange) predicted levels of reasoning about prosocial behaviour. Nevertheless, even though the value of differing opinions during adolescence and beyond can be hypothesized given such studies (and this is how it will be treated in the research to follow¹), the volume of relevant material remains lower than at younger ages. Thus, the hypothesis requires further testing, and the following research attempts to do this.

Whatever the case as regards differing opinions, evidence about collaborative constructions from adolescence onwards is extremely limited. A couple of studies hint at their irrelevance (i.e. Kapur, 2008; Sampson & Clark, 2009), and in any event it is hard to imagine the active processes accessed in childhood being lost at later age levels. Yet older groups engage more heavily in social practices where group decisions are treated as momentous, that is recorded in minutes and so on. For this reason, they might be expected to place greater emphasis upon collaborative constructions, and indeed the emphasis might even increase during adolescence. Furthermore, analyses of collaborative dialogue with adolescents and young adults, in contrast to work with children, often address 'rebuttal' and 'counter-argument' (e.g. Asterhan & Schwarz, 2009; Felton & Kuhn, 2001; von Aufschnaiter, Erduran, Osborne, & Simon, 2008), functions which, as conceptualized in these analyses, signify attempts to engage deeply with and resolve differences. This means that collaborative constructions are not merely pursued but also built on the very dialogue that, by hypothesis, is central to growth. Indeed, when rebuttal and counter-argument involve finding evidence that simultaneously supports one perspective and undermines others, a level of co-ordination is implicated that is rare prior to adolescence and only gradually emerges thereafter (Howe, Tolmie, & Sofroniou, 1998; Inhelder & Piaget, 1958; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995;

Schauble, 1990). This may be why Felton and Kuhn (2001) found ‘strategic sequences’ (as they call them) that include rebuttal and counter-evidence to be more frequent in their adult sample than amongst their young adolescents. In any event, through rebuttal and counter-argument too, there are hints of a relation between collaborative constructions and knowledge growth in adolescence and adulthood that is not simply stronger than in childhood but also strengthens with age.

The work of Howe et al. (1992a) and Tolmie and Howe (1993) might be interpreted as offering empirical support for both the relation and the putative age change. In the Howe et al. study, each pair of undergraduates was assigned a ‘group strategy score’ representing the quality of the explanatory factors that were mentioned. Group strategy scores showed a strong, positive association with pre- to post-test change. As detailed in Howe (2010), the discussions held by Tolmie and Howe’s teenagers contained features compatible with transition between non-reliance and reliance upon collaborative constructions. However, Howe et al.’s group strategy scores were derived from the interactions in general, not from collaborative constructions specifically. In addition, both group strategy scores and pre- to post-test change scores were associated with other variables, e.g. discussion of theoretical principles and of relevant data from earlier in or beyond the task, making directions of causality hard to decipher. This reflects the broader problem that neither this study nor Tolmie and Howe’s research were designed to address collaborative constructions, making attribution of relevance entirely post-hoc. More generally still, the fact remains that, as noted, close associations between collaborative constructions and individual growth do not *prove* that the former determine the latter: active decision-making remains possible. Should many studies point in the same direction and none suggest differently, a determining role might be presumed, but with two studies only nothing can be concluded. Recognizing this, the research that is reported below makes

a further attempt to examine the role of collaborative constructions during adolescence and beyond. Specifically, as well as testing the hypothesis that discussion of differing opinions remains relevant, it explores the possibility that collaborative constructions are also implicated. Moreover in view of the hints in previous research, it examines whether the relevance of collaborative constructions increases across the age range.

2. Method

The study involved students whose ages ranged from 10 years to early 20s. The students worked in pairs on a computer-presented task (hereafter the ‘collaborative task’) that required them to predict the trajectories objects follow when they fall from stationary or moving carriers, i.e. the topic used in Tolmie and Howe (1993 - see above). Extensive previous research indicates that students in this age range typically predict correct vertical fall from stationary carriers, but seldom anticipate parabolic fall in the direction of motion when carriers are moving (Anderson, Tolmie, Howe, Mayes, & Mackenzie, 1992; Eckstein & Kozhevnikov, 1997; Eckstein & Shemesh, 1989; Howe, Taylor Tavares, & Devine, 2012; Krist, 2000; Marioni, 1989; McCloskey, 1983; Whitaker, 1983). With moving carriers, even adults typically expect objects to fall vertically, travel backwards, fall diagonally forwards, or continue horizontally in space before making a 90-degree turn and falling. Thus, the topic was one that should challenge, and so provide potential for knowledge growth, throughout the study’s age range. In addition, the collaborative task was known from prior research to support learning in small-group contexts (Howe, Devine, & Taylor Tavares, 2013). In the present study, the students were recorded while completing the task, with growth assessed through change from pre-tests administered before working in pairs to post-tests administered some weeks afterwards. With possible age

differences considered throughout, analyses examined whether knowledge grew pre- to post-test, pair dialogue contributed to growth, and the quality of collaborative constructions was relevant. As detailed later, collaborative constructions were treated throughout as task-relevant conclusions that groups converged upon, while sometimes being operationalized with reference to computer input and sometimes with reference to dialogue.

2.1. Participants

Ninety students completed the pre-test, collaborative task and post-test. Twenty were drawn from Year 6 of a primary school located in rural East Anglia, England (Age range=10.58 to 11.50 years; $M=11.01$ years). A further 54 students were drawn from three year-groups of a single secondary school, this school also located in rural East Anglia. Twenty secondary students were in Year 8 (Age range=12.42 to 13.83 years; $M=12.97$ years), 20 were in Year 10 (Age range=14.25 to 15.08 years; $M=14.67$ years), and 14 were in Year 12 (Age range=16.25 to 17.58 years; $M=17.11$ years). The final 16 students were undergraduates at the University of Cambridge, studying Education or Medicine (Age range=18.33 to 22.00 years; $M=20.20$ years).

2.2. Materials

2.2.1. Collaborative task

As used in Howe et al. (2013), the task included ‘speed items’ that addressed speed change during fall as well as ‘direction items’ that addressed trajectory. The present study was restricted to the direction items, but otherwise used Howe et al.’s software without modification (so see the 2013 publication for details, including the rationale for task design). In brief, the software, which was programmed using Macromedia

Director, used eight direction items, all involving scenarios where a girl dropped a ball from a hot-air balloon. The scenarios depicted all possible combinations of whether: a) the balloon was stationary or moving at the time the ball was released; b) the ball fell onto grass or into a swimming pool; c) a green or silver ball was used, with the green ball shown via a real equivalent to be very light and the silver ball (of identical size) shown to be extremely heavy. The relevant combination was highlighted at the start of each scenario, using text and images (including motion from left to right when the balloon was moving). Scenario order was fixed, with stationary scenarios first, second, fifth and sixth.

At the moment the ball was released, the action froze and predictions of subsequent motion were invited. In particular, three white circles appeared under the balloon: a) directly below; b) behind; c) in front. This was accompanied with an instruction to click on the point that the ball would travel through. Once a circle was selected, this turned red and the other circles disappeared. At the same time, three further white circles appeared below the selected circle: a) if the selected circle was directly under the balloon, the new circles were directly below the selected one, behind and in front; b) if the selected circle was behind the balloon, the new circles were directly below, parabolically behind and diagonally behind; c) if the selected circle was in front of the balloon, the new circles were directly below, parabolically in front and diagonally in front. The instruction to select a point re-appeared. Once a second circle was chosen, this too turned red, the other circles disappeared and three further circles appeared below the second circle in the same relative positions as the second sets, accompanied with the instruction to select. The computer programme was designed to store all selections.

Correct predictions triggered the message ‘Well done! You are correct’ inside

a gold star to the right of the screen. This was accompanied with an invitation to see the ball fall, acceptance of which activated a simulation of the ball falling naturally. Incorrect predictions triggered an invitation to see what was anticipated, with acceptance activating a simulation of the ball falling non-naturally in the predicted fashion, and a request to indicate whether the motion looked correct. From Howe et al. (2012, 2013), it was expected that the non-naturalness would be recognized as such, even when it had been predicted. In any event, a message appeared subsequently confirming that the prediction had been wrong, together with invitations to see what really happens and interpret the difference between the incorrect and correct motion. The option was available of replaying the incorrect and/or correct motion as many times as was deemed necessary.

2.2.2. Pre- and Post-tests

Developed for the study reported in Howe et al. (2013), the pre- and post tests addressed both the prediction and explanation of object fall trajectories². The two tests comprised the same eight items, with each item associated with one of three scenarios: a) a ball falling from a hot-air balloon as with the collaborative task (3 items); b) a box falling from a helicopter (3 items); c) a box falling from a train on a high bridge (2 items). The first four items covered fall from rest, and the final four covered fall after motion. Each block of four covered all possible combinations of fall onto grass vs. into water and fall of a heavy vs. light object, with item order within blocks randomly determined. For presentation purposes, each test item comprised an illustration of the scenario accompanied with text.

The text included instructions to indicate where the objects would land by marking each illustration with a cross and then to draw the paths that the objects would follow as they fell. Four items also invited explanations of predicted paths,

offering six multiple-choice options (again in randomly varying orders) as possible factors: ‘A force pulling the ball/box downwards’, ‘The balloon/helicopter /train’s direction before the ball/box started falling’, ‘The air or water pushing the ball/box upwards’, ‘There is no wind’ (absence of wind was highlighted in the task instructions – see below), ‘The air or water pulling the ball/box downwards’, and ‘The ball/box’s weight’. The first two factors were regarded as relevant and the remaining four were regarded as irrelevant (but indicated in the background literature as often used when reasoning about this topic).

2.3. Procedure

With each of Years 6, 8, 10 and 12, the pre-test was presented in a single sitting to one full class of students; with the undergraduates, it was presented at mutually convenient times to small groups. A total of 118 students completed the pre-test, each working individually. Prior to presentation, a researcher gave each student a booklet, which contained the test items in sequence (i.e. illustrations and text), and invited them to insert background information on the front cover (e.g. name, gender, date of birth). Thereafter, she displayed the items in sequence on a large screen using PowerPoint, ensuring that the students 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, i.e. whether the carrier was stationary or moving prior to the ball/box’s release, whether the ball/box was light or heavy, and the intended absence of wind. Completion of the eight items took between 10 and 15 minutes.

The collaborative task was presented about one week after the pre-test. The goal had been to have 10 pairs at each of the five age levels work on the task, with the differing views which previous research pinpoints present within each pair. Given the

relative straightforwardness of comparing predicted paths (against, say, selected explanatory factors), difference was addressed with reference to paths. Accordingly, the eight paths drawn by each student were coded as ‘backward’ (i.e. in the opposite direction from the moving scenarios’ pre-fall motion), ‘vertical’, ‘forward non-parabolic’, ‘forward parabolic’, and ‘other’. Pairs were formulated such that members differed over at least five paths. Mixed-sex pairs were avoided given Tolmie and Howe’s (1993) evidence for their reluctance to discuss differences. Moreover, while not critical given the study’s aims, an attempt was made to balance the numbers of male-male and female-female pairs at each age level. In the event, five male-male pairs and five female-female pairs were formulated from the Year 6 and Year 8 samples, and six male-male pairs and four female-female pairs were formulated from the Year 10 sample. With fewer students available for pre-testing in the two oldest groups, only eight pairs were possible given the difference criterion. With the Year 12 students, five pairs were male-male and three were female-female; with the undergraduates it was precisely the reverse.

The collaborative task was presented via a Dell Latitude D820 laptop. With the software ready to use and an audio-recorder switched on (and pair identifiers recorded on both), the researcher who had administered the pre-test gave each pair an overview of the task, indicating how the scenarios would vary (using real balls to highlight the ball contrast) while also explaining that absence of wind should be assumed throughout. Then taking each of the task steps in turn, the researcher emphasized how important it was that every decision should be discussed thoroughly and agreement reached before inputting onscreen. Once she was convinced that everything was understood, she withdrew to another part of the room, only intervening when asked for procedural advice. The Year 6 pairs and undergraduates completed the collaborative task sequentially in a private room at their school or

university. At the school's request, the Year 8, 10 and 12 pairs completed the task in a large laboratory, with four or five pairs working simultaneously but placed out of earshot. Across the sample, the pairs took between 5 and 18 minutes to complete the task. The post-test was administered about four weeks after the collaborative task, following procedures that were identical to the pre-test. All pair members bar two were present for the post-test, meaning that, as noted, 90 students completed the procedure's three key stages.

2.4 Data preparation

2.4.1. Pre- and post-tests

To address the study's aims, it was necessary to assess pre- and post-test responses, code the dialogue that took place while the pairs worked together, and evaluate any collaborative constructions that emerged. As regards the pre- and post-tests, each test item was initially coded separately, and as indicated above coding of pre-test paths began prior to the collaborative task. To convert the qualitative codes into scores: a) each path predicted from a stationary carrier was scored 1 if it had been coded 'vertical' and 0 otherwise; b) each path predicted from a moving carrier was scored 2 if it had been coded 'forward parabolic', 1 if it had been coded 'forward non-parabolic', and 0 otherwise. In addition, a count was made of the number of relevant explanatory factors ticked with each item (maximum=2 per item) and the number of irrelevant factors (maximum=4 per item). Post-test paths and factors were coded likewise. Prior to full coding, the main coder and the first author independently coded 25% of the pre- and post-tests. They agreed regarding 97.41% of the pre-test items and 98.84% of the post-test items.

To examine how item scores might be combined, principal components

analysis with varimax rotation was conducted on respectively path scores, numbers of relevant factors, and numbers of irrelevant factors. Pre- and post-test scores were analysed separately. These analyses, like all those reported subsequently, were conducted using the *Statistical Package for the Social Sciences* Version 22 (SPSS, Chicago, IL, USA). With the path scores, a strong two-component solution emerged for both the pre- and post-tests. All items relating to stationary carriers loaded on one of these components (pre-test loadings=0.54 to 0.72; post-test loadings=0.61 to 0.80), and none of the items relating to moving carriers (pre-test loadings=0.01 to 0.11; post-test loadings=0.09 to 0.18). This component accounted for 21% of the variance at both pre- and post-test. All items relating to moving carriers loaded on the other component (pre-test loadings=0.86 to 0.91; post-test loadings=0.89 to 0.92), and none of the items relating to stationary carriers (pre-test loadings=-0.01 to 0.15; post-test loadings=0 to 0.12). This component accounted for 41% of the pre-test variance and 49% of the post-test variance. With the relevant and irrelevant factors, single component solutions emerged in both cases, with all four items loading on these components in every case (pre-test loadings for relevant (55% of variance)=0.66 to .80; post-test loadings for relevant (65% of variance)=0.76 to 0.83; pre-test loadings for irrelevant (59% of variance)=0.71 to 0.86; post-test loadings for irrelevant (70% of variance)=0.81 to 0.87). Based on these values, it seemed reasonable to use four measures of pre- and post-test performance as detailed in Table 1 (Pre- and post-test scores): paths predicted from stationary carriers (Path_S), paths predicted from moving carriers (Path_M), relevant explanatory factors (Expl_R), and irrelevant explanatory factors (Expl_I).

[Insert Table 1 about here]

2.4.2. *Group dialogue*

The audio-recordings were professionally transcribed, with the transcriber uninformed about the study's aims. The transcripts were then coded, with the coder unaware of both aims and student age. The coder ignored: a) comments from the researcher; b) comments about task procedures; c) reading from the screen; d) comments prior to the first item; e) off-task comments. Apart from this, all contributions were coded using the categories detailed in Table 2, these categories being designed to comply with the study's aims in light of the background literature. In particular, the categories reflected the fact that, as with any problem-solving task, the dialogue could focus upon task solutions (path categories, EP, DP, LP) or explanatory principles (factor categories, EF, DF, LF, and recalling Howe et al. (1992a), theory categories, ET, DT, LT). They acknowledged that factors could be used to identify situational parameters (IF) as well as explain (EF, DF, LF). For each of path, factor and theory, the categories also recognized the key distinction between proposing/accepting (endorse categories, EP, EF, ET) and differing (doubt categories, DP, DF, DT) and, again with reference to Howe et al. (1992a), they recorded references beyond the current problem (link categories, LP, LF, LT, LB).

[Insert Table 2 about here]

Normally, each speaker turn could be placed within a single category. Occasionally however, turns could have been coded using several categories, and here priorities were set: a) if path and factor categories could both apply, factor categories were preferred; b) if theory categories could apply together with path, factor or both, theory categories were preferred. Prior to full coding, the first author independently coded two transcripts from each age group, and computed the correlations between her category frequencies per task item and those obtained by the main coder.

Correlations varied between 0.66 and 0.96 depending on the category, with a mean of 0.82 across categories.

Upon completion of coding, it became apparent that the LT category was never used, and that the incidence of some other categories was very low when the task items were considered separately. Combination across items was required but given the clear distinction over pre- and post-test paths (see above), it was decided to separate the dialogue category totals for the four items relating to stationary carriers from the totals for the four items relating to moving carriers unless subsequent analysis recommended combination. Statistically significant positive correlations were obtained between the totals for stationary and moving carriers with IF, EP, LP, LF and ET ($df=45$, $p=0.02$ to < 0.001), so here combination did take place across the full eight items. This produced TotIF, TotEP and so on. With the other five categories the totals for the stationary and moving carriers were kept separate, producing TotDP_S (for stationary), TotDP_M (for moving) etc. With TotDT_S excluded since DT never occurred with stationary carriers, this resulted in 14 composite categories.

Next an attempt was made to condense the categories further, using standard reduction procedures (Field, 2013). First, correlations were computed between the frequencies of the 14 composites across the transcripts. TotDP_S and TotLB_S failed to correlate significantly with any other category, and so were eliminated from further analysis. They were in fact infrequent, for the remaining 12 categories jointly accounted for 98.41% of the coded dialogue. These 12 categories were then subject to principal components analysis with direct oblimin rotation. Oblique rotation was chosen because there was no a priori reason to think emergent components would be unrelated. Quality parameters were all acceptable, that is the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.75, Bartlett's test of sphericity was significant

($p < 0.001$), and values in the auto-image matrix were all greater than .5 (0.69 to 0.86). Thus, there was every reason to treat the three components that emerged with eigenvalues greater than one as the key dimensions of dialogue, and this was the strategy adopted for subsequent analyses. Accordingly, each transcript (and so each pair of students) was assigned three dialogue scores, representing the total frequencies within that transcript of the categories that, from component loadings, were associated with each of the three dimensions.

The dimensions are listed in Table 1 (Group dialogue). with the first one (from a component accounting for 35% of the total variance) associated with TotIF, TotEP, TotLP, and TotEF_S (respective component loadings=0.78, 0.77, 0.44, 0.81, with loadings for the other categories all between 0.29 and -0.25). Jointly, these categories accounted for 77% of the dialogue covered in the principal components analysis. An example of each of the four categories (sometimes one of several such examples) is highlighted in the sequence that follows from a Year 8 pair. Together, they amount to the non-contentious joint construction of paths, suggesting that the dimension could be termed ‘solution confirmation’:

Anna: [Reads - Notice which ball is used in the trial] The heavy one. [TotIF]

Holly: Splash. Right.

Anna: [Reads – Notice if the balloon moves] It’s not moving. [Reads – Select the next point the ball will travel through] So it’s the heavy one and it’s not moving.

Holly: It’s the heavy one so it will probably just go straight down. [TotEF_S]

Anna: It would go straight down wouldn’t it, with this bit? [TotEP]

Holly: Yeah.

Anna: Like it did with the other one, but we did it wrong. [TotLP]

The second dimension (from a component that accounted for 12% of the variance) was associated with TotDP_M, TotEF_M, TotDF_M, and TotLB_M (respective component loadings=0.71, 0.55, 0.71, 0.79, with other categories loading between 0.22 and -0.02). In other words, this dimension was exclusive to the moving scenarios, and involved divergence and negotiation around both paths and factors, suggesting the label 'solution/factor divergence'. Jointly the four constituent categories accounted for 16% of the dialogue included in the principal components analysis. With examples of the categories highlighted, solution/factor divergence can be illustrated via the following sequence from undergraduates:

Sonia: That's what I personally think because once she's released the ball, her movement is not going to have any effect. *[TotEF_M]*

Naomi: I don't know. I seem to remember that like if you go this way, it falls, that the direction makes a difference. *[TotDF_M]*

Sonia: Like maybe because you're thinking when you're moving – like are you thinking that you drop something out of a car and it starts moving?
[TotLB_M]

Naomi: Yes exactly.

Sonia: You're thinking it's moving in the background but actually it's just falling down. It'll go straight down.

Naomi: But you see it rolling back, so it'll go back. *[TotDP_M]*

Sonia: I don't know.

The final dimension (whose associated component accounted for 16% of the variance) encompassed TotET, TotDT_M, TotDF_S, and TotLF (respective component loadings=0.89, 0.69, 0.88, 0.65, with the loadings of other categories ranging from 0.23 to -0.17). Jointly, these categories covered 6% of the dialogue included in the principal components analysis. When one of the dimension's

constituents was specific to moving scenarios and another was specific to stationary scenarios, no short sequence can illustrate every feature. However, the following exchanges between undergraduates underline how theory contributed to the dimension, and how like the second dimension it too involved divergence and negotiation. Thus, the label ‘theory divergence’ seems appropriate:

Vanessa: Probably because the velocity gradually decreases, like in the horizontal one because, I don’t know, I’m thinking about physics. [*TotET*]

Lindsay: I’m not thinking about it now. I’m just thinking I don’t know because it wasn’t going that fast, it didn’t have that much momentum behind it.

Vanessa: Yeah.

Lindsay: What do you think?

Vanessa: Actually, I’m thinking like about physics, like if there’s a horizontal component and a vertical component and the resolved one is actually that plus that one, so it’s actually, it should be like a straight slope.

Lindsay: You think so? [*TotDT_M*]

Vanessa: But in that case, I don’t know why, it’s like if there is no air friction, then the horizontal component should be the same all the time, so that’s why it should be like a straight line instead of a curved one. But this one is curved.

Lindsay: Shall we just say we’ve thought about it?

2.4.3. Collaborative constructions

If individual pre- to post-test progress depends upon the appropriation of collaborative constructions, there will be group conclusions that parallel every pre- to post-test measure on which progress is detected, and the quality of these conclusions will be positively associated with the extent of progress. The implication is that to test this conjecture, four indices of collaborative construction are needed, corresponding to the

four pre- and post-test measures. With what might be called CoPath_S and CoPath_M (i.e. collaboratively constructed paths from respectively the stationary and moving carriers), there was actually a choice: the indices could be operationalized as the number of correct paths inputted onscreen, or as the number of correct paths mutually endorsed in dialogue, i.e. proposed by one student and accepted by their partner. As indicated in Table 1 (Collaborative constructions), the first approach was preferred: a) it meant that the indices could be computed with total reliability, from the stored computer input if necessary but even from the transcripts; b) it optimized the assessment of quality. As regards the latter, the pairs inputted an average of 3.70 correct paths across the four stationary scenarios, but only stated the correct paths on an average of 3.33 of these scenarios and indicated mutual endorsement on an average of 2.76. The corresponding averages for the four moving scenarios were 0.74 for correct input, 0.56 for correct statements, and 0.46 for mutual endorsements.

With CoExpl_R and CoExpl_I (i.e. collaboratively constructed accounts of respectively which explanatory factors were relevant and which were irrelevant), there was no alternative but to utilize dialogue, and within dialogue there was no task requirement that explanations should be referred to, and no fixed number of opportunities for reference. Thus to obtain indices, every instance was identified where one member of each pair used a factor correctly in an explanatory sense (i.e. their turn was coded EF, DF or LF and the F was correct), and their partner showed that they endorsed this usage. Identified instances were then divided into those where relevant factors were collaboratively accepted, e.g. 'It's not moving so it will go straight down', 'Yes, that's right, it's only when it's moving that it goes to one side', and those where irrelevant factors were collaboratively rejected, e.g. 'I don't think it matters if it's water or anything to be honest', 'No it doesn't matter'. Frequency counts of these sub-types produced CoExpl_R and CoExpl_I respectively.

3. Results

Investigating the role of group dialogue and collaborative construction in individual growth presupposes that individual understanding did in fact grow. Accordingly, the first step in the analysis involved examining pre- to post-test change in Path_S, Path_M, Expl_R and Expl_I, via four two-way mixed model ANOVAs³ with Test (Pre-test score, Post-test score) as the repeated measure and Age (Year 6, Year 8, Year 10, Year 12, Undergraduate) as the between-groups measure⁴. With Path_S, there was a statistically significant main effect of Test, $F(1, 85)=25.16, p<0.001$, partial $\eta^2=0.23$: as Table 3 shows, the students progressed between pre- and post-test. There was also a significant main effect of Age, $F(4, 85)=5.28, p=0.001$, partial $\eta^2=0.20$. Moreover, while the Test x Age interaction was not statistically significant⁵ post-hoc tests revealed some differences between pre- and post-tests in how the age groups performed. As indicated in Table 3, there were no significant age differences at pre-test, but at post-test the Year 8, Year 12 and undergraduate students outperformed the Year 10 students with the Year 6 students lying in-between. Path_M was also associated with statistically significant main effects of Test and Age, respectively $F(1, 85)=79.05, p<.001$, partial $\eta^2=.48$ and $F(4, 85)=19.90, p<.001$, partial $\eta^2=.48$. However, this time the Test x Age interaction was also significant, $F(4, 85)=11.20, p<0.001$, partial $\eta^2=0.35$. As can be seen in Table 3, there were no differences between the age groups at pre-test, but marked differences at post-test. The Year 10 students failed to progress, and while progress was discernible in all other groups it was especially marked amongst the Year 12 students and the undergraduates.

[Insert Table 3 about here]

With Expl_R, there was no statistically significant main effect of Test and no statistically significant Test x Age interaction. Thus with this measure, the students did not progress from pre- to post-test. Moreover, although the main effect of Age was statistically significant, $F(4, 85)=4.83$, $p=0.001$, partial $\eta^2=0.19$, post-hoc tests ($p<.05$) detected no significant differences between any groups (see Table 3). Finally, with Expl_I, there were statistically significant main effects of Test and Age, respectively $F(1, 85)=5.78$, $p=0.02$, partial $\eta^2=0.06$, and $F(4, 85)=3.80$, $p=0.007$, partial $\eta^2=0.15$. As Table 3 indicates, there were no significant differences between the age groups at pre-test, but while four groups showed the decrease in mean scores that, with this measure, signifies progress, the magnitude of the decrease varied. This said, the Test x Age interaction was not statistically significant. Overall then, pre- to post-test growth occurred with Path_S, Path_M and Expl_I warranting their use in subsequent analyses. However, there was no progress with Expl_R, which meant that this measure had to be dropped.

3.1. Group dialogue

Two further aspects of pre- to post-test change are worthy of note. First, despite the broad age range and contrasting educational experiences, there were no significant differences between the age groups over pre-test scores. This indicates that the task did not merely prove challenging at all age levels as hoped; it also offered challenges that were more-or-less equivalent across the groups. Second, there were marked age differences over post-test scores with Path_M and Expl_I, and some differences with Path_S. Since all pairs inputted eight paths and received feedback on accuracy, this suggests that any role which the computer feedback played cannot have been straightforward or sufficient. The pointers are very much towards a contribution from dialogue, especially when as Table 4 shows the age profiles for two dialogue

dimensions are roughly in line with knowledge gain. In particular, the frequencies of both solution/factor and theory divergence were lowest in Year 10, and the Year 10 students showed the least pre- to post-test gain. This said, when the mean frequencies were compared as a function of age using one-way ANOVAs, only solution confirmation was associated with statistically significant differences, $F(4, 41)=4.10$, $p=0.007$. As Table 4 indicates, solution confirmation was especially frequent amongst the two youngest groups.

[Insert Table 4 about here]

To examine the relation between dialogue and knowledge growth (and, as a second step, to see whether the relation changed with student age), each student was assigned three scores, corresponding to the frequency with which dialogue associated with the three dimensions occurred within their pair⁶. Multiple regression was then used to ascertain whether any of the dimensions predicted post-test score. As shown in Table 5, dialogue was only weakly associated with post-test Path_S score, and of the dimensions only solution/factor divergence proved to be a statistically significant predictor. This is despite the fact that solution/factor divergence only occurred when the carrier was moving. With Path_M, all three of the dialogue dimensions were significantly associated with post-test score. However, while solution/factor and theory divergence were positively predictive of growth, solution confirmation was negatively predictive. A similar pattern emerges with Expl_I once it is remembered that here low post-test scores are indicative of growth. Therefore, the negative beta values for solution/factor and theory divergence signify positive prediction of growth, with the value statistically significant for theory divergence. Again though, the more frequently the students engaged in solution confirmation the less they progressed.

[Insert Table 5 about here]

To examine potential variation with student age, the three dialogue dimensions were centred, the students' ages in months were calculated and also centred, and the centred Age x Dimension interactions were computed. Hierarchical multiple regressions were conducted with the dialogue dimensions entered at the first level, and age and the interactions entered at the second level. Age effects replicated those reported earlier, none of the interactions approached statistical significance, and as regards dialogue the picture presented in Table 5 was endorsed. In general then, when dialogue was implicated in growth it was solution/factor and theory divergence that emerged as positively predictive and this was true regardless of age. This supports the hypothesis that discussion around differing opinions is as relevant in adolescence and early adulthood as it is with younger groups.

3.2. Collaborative constructions

With dialogue established as predictive of growth, it becomes appropriate to examine whether the relation depended upon the quality of collaborative constructions. Given the focus upon Path_S, Path_M and Expl_I, three indices of quality are relevant, CoPath_S, CoPath_M and CoExpl_I. As indicated in Table 6, mean CoPath_S scores were always close to the theoretical maximum of 4.00, implying that the groups mainly converged on the correct paths when the balloon was stationary. There were no significant age differences. Also with a theoretical maximum of 4.00, mean CoPath_M scores were much lower, suggesting that despite the correct paths being simulated the moving carriers continued to challenge. Although the undergraduates achieved a higher mean score than the other groups, due to improved performance with the final two problems, age differences once more were not statistically significant. As for CoExpl_I scores, they were consistently low, indicating that

irrelevant factors were seldom ruled out during the group discussion. Here too, there were no significant differences as a function of age.

[Insert Table 6 about here]

Given the strong age differences over pre- to post-test growth, it is difficult given Table 6 to envisage a major role for collaborative constructions. Nonetheless, an attempt was made to examine the role directly, recognizing that the notion that their quality determines the impact of dialogue constitutes a ‘mediational’ hypothesis (Baron & Kenny, 1986). Therefore it can be examined using standard techniques for testing mediation (see, e.g., Field, 2013). Applying the techniques to the present data and noting the interest not simply in whether mediation occurs but also in whether its extent changes with age, the preliminary steps involved testing whether: a) aspects of dialogue predict post-test scores with age differences considered, i.e. what was reported above; b) quality indices relating to collaborative constructions predict post-test scores with age differences considered; c) relevant aspects of dialogue predict quality indices, again with age considered.

Focusing first on the relation between the quality of collaborative constructions and post-test performance (i.e. b) above), three multiple regression analyses were conducted, one with CoPath_S, Age and the Age x CoPath_S interaction as predictors (all variables centred) and post-test Path_S as the dependent variable and the other two equivalent but using CoPath_M/Path_M and CoExpl_I/Expl_I. CoPath_S did prove to be a statistically significant predictor of post-test Path_S, $\text{Beta}=-0.23$, $t=2.04$, $p=0.04^7$. However, neither CoPath_M nor CoExpl_I were significantly associated with the corresponding post-test scores, respective beta values being -0.14 , $t=1.16$, $p=0.25$ and 0.48 , $t=1.12$, $p=0.28$. With all three analyses, age effects confirm those reported earlier, and the interaction effects

were not statistically significant. Thus, with Path_M and Expl_I, mediation can be rejected at all age levels without further analysis, meaning that collaborative constructions were not implicated in the strong growth that was detected with these variables. By contrast with solution/factor divergence and now CoPath_S established as predictive of growth, further analyses were warranted with Path_S. Accordingly, regression was used to examine whether solution/factor divergence predicted CoPath_S, again with possible age differences considered (i.e. c) above). The effect of divergence was statistically significant, $Beta=-0.35$, $t=3.34$, $p=0.001$, once more without associated interaction with age.

The triangular relation amongst solution/factor divergence, CoPath_S score and post-test Path_S score that the above implies means that with the paths from stationary carriers, the preconditions for mediation were fulfilled. Mediation per se was tested following Preacher and Hayes' (2004) bootstrapping procedures, no longer including interaction terms when no significant interactions with age were detected during the preliminaries. The results are shown in Figure 1. Consistent with mediation, the indirect effect of solution/factor divergence on Path_S through CoPath_S was statistically non-significant. Yet the confidence intervals indicate that mediation via CoPath_S cannot reliably be attributed.

[Insert Figure 1 about here]

4. Discussion

The starting place for the study was research into small-group collaboration during middle to late childhood. This research shows that while individual knowledge and understanding can be promoted through exchanging different opinions, the joint analyses that groups construct while collaborating play a tangential role. Individuals

may or may not accept these analyses depending upon processes of reflection and reconciliation that are triggered through difference and sometimes occur post-group. Noting a relative dearth of research with older participants, the study considered whether equivalent processes operate during adolescence and early adulthood. In the event, exchanges around differing opinions made an unmistakable contribution at this level too. As was clear from Table 5, solution/factor and theory divergence were both strongly and positively predictive of progress with the paths drawn from moving carriers (Path_M), and one of these variables was also strongly and positively predictive of progress with the paths drawn from stationary carriers (Path_S) and the elimination of irrelevant explanatory factors (Expl_I). By contrast, the one dialogue dimension that did not involve difference, solution confirmation, showed a consistently negative relation with progress. This relation was highly significant with the paths drawn from moving carriers and the use of irrelevant factors.

These associations between dialogue and progress were constant across the age range, for none of the Age x Dialogue interactions were significantly associated with post-test score. The implication is that when progress varied with age, it was a straightforward consequence of varying dialogue frequencies. Thus, the reason why the Year 6 and 8 students progressed less than the Year 12 students and the undergraduates with paths from moving carriers and irrelevant factors was that, as Table 4 indicates, they produced less theory divergence and more solution confirmation. The four groups were very similar over solution/factor divergence. The variation over theory divergence most likely results from education: by Year 12, students will typically have covered the relevant physics, and even if key principles are insecurely grasped (as with the third of the dialogue sequences quoted earlier) they will provide repertoires of constructs to be fed into discussions. The variation over solution confirmation stemmed from more marked decreases in frequency as the

task progressed amongst the older groups: perhaps these groups more rapidly developed shared understanding of what to input even when, as often happened with the moving scenarios, the input was wrong. Likewise, the reason why the Year 10 students were the least successful of all groups was that they produced solution/factor and theory divergence with the lowest frequencies. Earlier, reference was made to Tolmie and Howe's (1993) evidence for reluctance to differ at this age level amongst mixed-sex pairs. The present pairs were all single-sex, but possibly there is general self-consciousness mid-teen (and desire to be 'cool') that works against the depth of analysis that exploring differences requires. Certainly, it was only the Year 10 students, who via comments like 'Are they done already?', showed sensitivity to groups elsewhere in the room.

With the paths drawn from moving carriers and the use of irrelevant factors, solution/factor and/or theory divergence were the only predictors of growth, for in neither case did collaborative constructions contribute. As shown in Table 6, the paths that were constructed jointly when the balloon was moving (CoPath_M) were consistently poor, and the instances where irrelevant factors were jointly rejected (CoExpl_I) were exceedingly rare. The modest variation over scores was unrelated to post-test performance. Thus, the paths from moving carriers and the use of irrelevancies provide clear-cut evidence during adolescence and beyond for precisely the processes identified with children, individual reflection and reconciliation triggered through difference. Learning via the appropriation of collaborative constructions simply did not occur. With the paths drawn from stationary carriers, the situation was more ambiguous: while the preconditions were fulfilled for the paths that were constructed jointly (CoPath_S) mediating the relation between solution/factor divergence and post-test score, formal analysis indicated that mediation cannot be confidently attributed. What is clear though is that, whether

mediated by collaborative constructions or not, post-test Path_S scores were predicted by solution/factor divergence, and this dimension of dialogue was restricted to the moving scenarios. It did not occur with the stationary scenarios to which CoPath_S and Path_S applied. The second of the extracts quoted earlier may help to explain this apparent paradox, for it illustrates how, as happened frequently, the paths that were proposed and contested when the balloon was moving included vertical fall. Framed as an implicit stationary/moving contrast, it may have heightened awareness of what was anticipated with stationary carriers, with knock-on effects for the subsequent stationary scenarios.

The notion that difference heightens awareness and this plays a key role in triggering growth is consistent with the three theorists discussed above, Piaget, Vygotsky and Bakhtin. In Piaget (1959) argument and opposition are depicted as important because they press each individual ‘to justify himself in the eyes of others and thus acquire the habit of watching himself think’ (p.137). For Vygotsky (1998) their crucial function is to confront the individual with ‘the need to form a basis, to prove, confirm and verify his own idea’ (p.168). In Bakhtin (1981), struggling with another’s discourse is regarded as the key to ‘ideological consciousness’ (p.348) and through this greater understanding within the one who struggles. Using contemporary terms, the notion is that difference during small-group collaboration stimulates a ‘meta-cognitive’ perspective upon relevant personal beliefs, and this was implicated in the positive consequences. Evidence already exists that collaboration can support meta-cognition including with computer-based tasks (e.g. Winne, Hadwin, & Perry, 2013), and given recent research, the suggestion that meta-cognition is implicated in growth is scarcely controversial: regardless of age, meta-cognitive perspectives have been shown repeatedly to be challenging to adopt (e.g. Kuhn, 1989, 1991) but once

adopted to be strongly supportive of knowledge gain (Hattie, 2008; Higgins, 2013; Van der Stel & Veenman, 2014).

More specifically, Kuhn (1989, 1991) shows that once meta-cognitive perspectives are adopted, they support the productive use of evidence as feedback on beliefs, and this is also the message from three studies reported by Howe and colleagues (Howe et al., 2000, 2005; Howe & Tolmie, 2003) relating to small-group collaboration per se. In different ways, all three studies demonstrate that when (and only when) beliefs were thrust into consciousness through discussion of differences, subsequent empirical evidence was treated as significant. Moreover, the studies also show that by virtue of being treated as significant, the evidence made a material contribution to eventual growth. Howe and colleagues' research was conducted with pre-adolescents, but the continuities that the present study has already demonstrated with younger groups suggest relevance here too. Certainly, responses to the computer feedback indicate that empirical evidence was treated as significant, for instance 'How did? What? Why would it go that way though?' 'I've no idea. It's literally the exact opposite'. In addition, Howe and colleagues' work also indicates that once a meta-cognitive perspective is adopted, it supports individual reflection, reconciliation, and productive reference to evidence for weeks after the triggering group work. This too may be relevant here, for with paths from moving carriers at least, there are strong suggestions of post-group processing: scores for collaborative constructions (CoPath_M) typically remained low even on the final problem suggesting little ontask progress, and as noted earlier the students were even less likely to articulate correct answers while collaborating than to input these onscreen.

While the above attributes a key role to the computer feedback, it is crucial to reiterate that the progress observed in the present study cannot have been a direct

response to this feedback: as highlighted earlier the feedback was constant across age groups when the amount of progress varied considerably. Rather, the proposal is that the progress was a response to feedback that was dependent upon a meta-cognitive orientation to personal beliefs that was in turn dependent upon differences expressed in dialogue. With the moving scenarios, the rational response to feedback amongst the study's sample would have been to reject the collaborative constructions as a guide to progress and to look elsewhere, since these constructions would have been even more transparently invalidated through the feedback than the beliefs that underpinned them. Thus, the proposal is entirely consistent with the lack of relation between the collaborative constructions of correct paths with the moving scenarios and post-test performance. With the stationary scenarios, initial understanding as revealed in pre-test scores was good, and the collaborative constructions were usually accurate. Thus, here the rational response to feedback would have been to regard personal beliefs as endorsed and in little need of refinement. Once more, collaborative constructions should generally have been treated as peripheral, although they might in some cases have also boosted confidence in what was presumed. This too is consistent with the study's results, for the ambiguities highlighted above around CoPath_S's mediating role are entirely compatible with an impact that varied from weak to non-existent. There is however another possible outcome from small-group collaboration that was not observed in the present study, namely students whose initial understanding was poor (as with the moving scenarios) collaboratively constructing reasonable solutions (as with the stationary scenarios). Here the rational response would be to rely heavily on collaborative constructions when reflecting upon how personal beliefs should be modified. Earlier, studies with children were cited where collaborative constructions were predictive of growth. Could it be that these studies created the missing combination? It would be a strong test of the present proposal to explore this further.

Reflecting on a relationship that did not occur underlines the fact that the study was not intended to deny contributions from collaborative constructions. Rather, the intention was to examine whether, during adolescence and early adulthood, progress after small-group collaboration *depends* upon progressive collaborative constructions emerging from discussion of differences because these constructions are the *source* of progress. This is what Vygotsky would have expected. In the event, the study confirms the relevance of differing opinions: pre- to post-test progress was detected with three of its four measures, and in all three cases dialogue around differences was implicated. However, it also demonstrates conclusively that progressive collaborative constructions are not essential: with the paths drawn from moving carriers and the elimination of irrelevant factors, these constructions were respectively poor and exceedingly rare, yet substantial pre- to post-test progress was detected. The signs are that the progress occurred post-group, albeit stimulated through dialogue. Thus, a process model that acknowledges individual reflection and reconciliation is certainly required for two of the four measures, and less obviously it is also needed for the paths drawn from stationary carriers. Here collaborative constructions may have made a modest contribution (although this is unproven). However, when the dialogue that stimulated growth, solution/factor divergence, did not even occur with stationary carriers, any use of collaborative constructions must here too have depended on active individual processing. A model that moves from difference through meta-cognition and feedback to knowledge growth fits these data, just as it fits and suggests continuity with data previously obtained from children.

4.1. Limitations

At present, the proposed model is post hoc, and for that reason alone requires further testing. Moreover, even if support is obtained additional issues will arise, which were

beyond the scope of the reported research. What for instance are the implications for topics that do not lend themselves to eye-catching, physical feedback? The identification of suitable topics would not be straightforward, given that, like trajectories during object fall, they would need to be challenging across a broad age range. Nevertheless, contrastive analysis across several topics would be desirable. A second issue relates to the precise nature of any association between dialogue and meta-cognition, specifically can appropriate meta-cognitive perspectives be promoted in alternative dialogic contexts or do they require small-group interaction amongst peers, can these perspectives be promoted through other forms of dialogue or are differences of opinion essential, and are the perspectives inconceivable in the absence of dialogue or could strategies be found to support, say, solitary learners? Here too there is a need for evidence that goes beyond what the present study could provide.

4.2. Conclusions

Supplementing the limited evidence relating to older participants, the reported study shows that discussion of differing opinions during small-group collaboration is productive for adolescents and young adults, just as previous research has shown this to be productive for children. Discussion of divergent views about task solutions, determining factors, and underlying theories were all found to be helpful. At the same time (and also consistent with research relating to children), the study demonstrates that productive outcomes were not dependent on progressive solutions being achieved while collaborating, i.e. upon high quality collaborative constructions. The substantial advances that were observed on two of the study's key measures were independent of such constructions, which in any event were seldom achieved. Qualifying the Vygotskian perspective upon growth after small-group collaboration, these results suggest a model where differences expressed in dialogue stimulate reflective, meta-

cognitive orientations towards personal beliefs, which support productive responses to feedback. The absence of age differences within the study's sample and the parallels with what is known about children suggest that the model may have applicability across a wide age range. While the model requires further testing, a contribution has hopefully been made already to what earlier was flagged as the study's broadest aim, clarification of the processes by which small-group dialogue triggers growth.

Footnotes

¹ Hypothesized effects were subject to one-tailed testing; all other reported tests are two-tailed.

² In the pre-test only, items addressing general reasoning skills followed the object fall items. Responses to the reasoning skill items would have been examined (and presented in a subsequent publication) had the results reported here turned out differently. Established from the present results to be redundant, the items are not described here for the sake of brevity.

³ Multiple ANOVAs are conventionally preferred over MANOVA when, as here, there are no grounds for presuming that dependent variables are linked (Field, 2013).

⁴ Key analyses were repeated with gender considered. With no significant Age x Gender interactions, gender effects are omitted in the interests of brevity.

⁵ Also for brevity, details of non-significant results are usually omitted, but are available upon request to the first author.

⁶ Multi-level analyses are often used when, as here, students are clustered in groups. However, for appropriate usage ‘treatments’ need to be extrinsic to the groups while here the main analyses, i.e. those reported in 3.1 and 3.2, address treatments that are intrinsic, namely group dialogue and collaborative constructions. Thus, the explored relations are two-level, rendering multi-level analysis inappropriate.

⁷ With centring, negative beta values mean positive prediction.

References

- Anderson, A., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science, 29*, 1-32.
- Anderson, A., Tolmie, A., Howe, C., Mayes, J., & 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.
- Asterhan, C., & Schwarz, B. (2009). The role of argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science, 33*, 373-399.
- Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development, 2*, 202-221.
- Bakhtin, M. (1981). *The dialogic imagination: Four essays*. Austin: University of Texas Press.
- Barnes, D., & Todd, F. (1977). *Communication and learning in small groups*. London: Routledge and Kegan Paul.
- Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Berkowitz, M., Gibbs, J., & Broughton, J. (1980). The relation of moral judgment stage disparity to developmental effects of peer dialogues. *Merrill-Palmer Quarterly, 26*, 341-357.
- Chi, M.T.H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.

- Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly*, 28, 347-367.
- Damon, W., & Phelps, E. (1988). Strategic uses of peer learning in children's education. In T. Berndt & G. Ladd (Eds.), *Peer relations in child development* (pp.135-157). New York: John Wiley & Sons.
- Doise, W., & Mackie, D. (1981). On the social nature of cognition. In J. Forgas (Ed.), *Social cognition: Perspectives on everyday understanding* (pp. 53-83). London: Academic Press.
- Doise, W., & Mugny, G. (1979). Individual and collective conflicts of centrations in cognitive development. *European Journal of Psychology*, 9, 105-108.
- Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford: Pergamon.
- Eckstein, S., & Kozhevnikov, M. (1997). Parallelism in the development of children's ideas and the historical development of projectile motion theories. *International Journal of Science Education*, 19, 1057-1073.
- Eckstein, S., & Shemesh, M. (1989). Development of children's ideas on motion: Intuition vs. logical thinking. *International Journal of Science Education*, 25, 671-688.
- Felton, M., & Kuhn, D. (2001). The development of argumentive discourse skills. *Discourse Processes*, 32, 135-153.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. Sage: Los Angeles.
- Fung, D., & Howe, C. (2014). Group work and the learning of critical thinking in the Hong Kong secondary liberal studies curriculum. *Cambridge Journal of Education*, 44, 245-270.
- Hattie, J.A. (2008). *Visible learning*. London: Routledge.

- Higgins, S. (2013). Self-regulation and learning: Evidence from meta-analysis and from classrooms. *British Journal of Educational Psychology Monograph Series II, 10*, 111-126.
- Howe, C. (2009). Collaborative group work in middle childhood: Joint construction, unresolved contradiction and the growth of knowledge. *Human Development, 39*, 71-94.
- Howe, C. (2010). Peer dialogue and cognitive development: A two-way relationship? In K. Littleton, & C. Howe (Eds.), *Educational dialogues: Understanding and promoting productive interaction* (Pp. 32-47). London: Routledge.
- Howe, C., McWilliam, D., & Cross, G. (2005). Chance favours only the prepared mind: Incubation and the delayed effects of peer collaboration. *British Journal of Psychology, 96*, 67-93.
- Howe, C., & Mercer, N. (2007). *Children's social development, peer interaction and classroom learning. The Primary Review (Research Survey 2/1b)*. Cambridge: University of Cambridge.
- 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., Tavares Taylor, J., & Devine, A. (2012). Everyday conceptions of object fall: Explicit and tacit understanding in middle childhood. *Journal of Experimental Child Psychology, 111*, 351-366.
- Howe, C., Devine, A., & Tavares Taylor, J. (2013). Supporting conceptual change in school science: A possible role for tacit understanding. *International Journal of Science Education, 35*, 864-883.
- 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., Anderson, A., & Mackenzie, M. (1992a). Conceptual knowledge in physics: The role of group interaction in computer-supported teaching. *Learning and Instruction*, 2, 161-183.

Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattray, C. (2000). Hypothesis testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction*, 10, 361-391.

Howe, C., Tolmie, A., & Rodgers, C. (1992b). The acquisition of conceptual knowledge in science by primary school children: Group interaction and the understanding of motion down an incline. *British Journal of Developmental Psychology*, 10, 113-130.

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. and Piaget, J. (1958). *The growth of logical thinking*, New York: Basic Books.

Jurkowski, S., & Hänze, M. (2015). How to increase the benefits of cooperation: Effects of training in transactive communication on cooperative learning. *British Journal of Educational Psychology*, 85, 357-371.

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26, 379-424.

Krist, H. (2000). Development of naïve beliefs about moving objects: The straight-down belief in action. *Cognitive Development*, 15, 281-308.

Kruger, A. (1992). The effect of peer and adult-child transactive discussions on moral reasoning. *Merrill-Palmer Quarterly*, 38, 191-211.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674-689.

- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995) Strategies in knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*, 245.
- Larrain, A., Freire, P., & Howe, C. (2014). Science teaching and argumentation: One-sided versus dialectical argumentation in Chilean middle-school science lessons. *International Journal of Science Education, 36*, 1017-1036.
- Leman, P., & Duveen, G. (1999). Representations of authority and children's moral reasoning. *European Journal of Social Psychology, 29*, 557-575.
- Littleton, K., & Mercer, N. (2012). Communication, collaboration, and creativity: How musicians negotiate a collective 'sound'. In D. Hargreaves, D. Miell, & R. MacDonald (Eds.), *Musical imaginations. Multidisciplinary perspectives on creativity, performance and perception* (pp. 233–241). New York: Oxford University Press.
- Marioni, C. (1989). Aspects of students' understanding in classroom settings (age 10-17): Case study on motion and inertia. *Physics Education, 24*, 273-277.
- McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*, London: Routledge.
- Miell, D., & Littleton, K. (2004). *Collaborative creativity*. London: Free Association Books.
- Miell, D., & MacDonald, R. (2000). Children's creative collaborations: The importance of friendship when working together on a musical composition. *Social Development, 9*, 348-369.

- Mollard, D. (2009). Moral development, social interaction, gender and the co-ordination of perspectives. Unpublished doctoral dissertation, University of Cambridge, UK.
- Mugny, G., & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology*, 8, 181-192.
- Piaget, J. (1959). *Judgment and reasoning in the child*. Michigan: Littlefield Adams.
- Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago: University of Chicago Press.
- Preacher, K.J., & Hayes, A.F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers*, 36, 717-731.
- Roseth, C.J., Johnson, D.W., & Johnson, R.T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin*, 134, 223-246.
- Roy, A., & Howe, C. (1990). Effects of cognitive conflict, socio-cognitive conflict and imitation on children's socio-legal thinking. *European Journal of Social Psychology*, 20, 241-252.
- Sampson, V., & Clark, D. (2008). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, 93, 448-484.
- Schauble, L. (1990). Belief revision in children: the role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49, 31-57.
- Schwarz, B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right....If they argue together! *Cognition and Instruction*, 18, 461-494.

- Silverman, I., & Geiringer, E. (1973). Dyadic interaction and conservation induction: A test of Piaget's equilibration mode. *Child Development, 44*, 815-820.
- Slavin, R.E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*, 427-515.
- Tolmie, A., & Howe, C. (1993). Gender and dialogue in secondary school physics. *Gender and Education, 5*, 191-209.
- Tolmie, A., Howe, C., Mackenzie, M., & Greer, K. (1993). Task design as an influence on dialogue and learning: Primary school group work with object flotation. *Social Development, 2*, 183-201.
- Van der Stel, M., & Veenman, M. V. (2014). Metacognitive skills and intellectual ability of young adolescents: A longitudinal study from a developmental perspective. *European Journal of Psychology of Education, 29*, 117–137.
- Vogel, F., Kollar, I., Ufer, S., Reichersdorfer, E., Reiss, K., & Fischer, F. (2016). Developing argumentation skills in mathematics through computer-supported collaborative learning: The role of transactivity. *Instructional Science, 44*, 477–500.
- Von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching, 45*, 101-131.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. (1998). *The collected works of L. S. Vygotsky. Volume 5: Child psychology*. New York: Plenum Press.
- Whitaker, R. (1983). Aristotle is not dead: Student understanding of trajectory motion. *American Journal of Physics, 51*, 352-358.

Williams, J., & Tolmie, A. (2000). Conceptual change in biology: Group interaction and the understanding of inheritance. *British Journal of Developmental Psychology, 18*, 625-649.

Winne, P. H., Hadwin, A. F., & Perry, N. E. (2013). Metacognition and computer-supported collaborative learning. In C. Hmelo-Silver, C. Chinn, C. K. Chan, & A. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 462–479). New York: Taylor & Francis.

Table 1

Variables Used in Analyses

Label	Definition
Pre- and post-test scores	
Path_S	Paths predicted from stationary carriers: Possible range per test=0 to 4, with four items each scoring 0 or 1
Path_M	Paths predicted from moving carriers: Possible range per test=0 to 8, with four items each scoring 0 to 2
Expl_R	Relevant explanatory factors: Possible range per test=0 to 8, with four items each scoring 0 to 2
Expl_I	Irrelevant explanatory factors: Possible range per test=0 to 16, with four items each scoring 0 to 4
Group dialogue	
Solution confirmation	Combined total frequency of <i>Identify factor</i> (TotIF) + <i>Endorse path</i> (TotEP) + <i>Link path</i> (TotLP), all across all scenarios + <i>Endorse factor</i> for stationary scenarios only (TotEF_S)
Solution/factor divergence	Combined total frequency for moving scenarios only of <i>Doubt path</i> (TotDP_M) + <i>Endorse factor</i> (TotEF_M) + <i>Doubt factor</i> (TotDF_M) + <i>Link beyond</i> (TotLB_M)

Table 1 (continued)

Label	Definition
Theory divergence	Combined total frequency of <i>Endorse theory</i> (TotET) + <i>Link factor</i> (TotLF), both across all scenarios + <i>Doubt theory</i> for moving scenarios only (TotDT_M) + <i>Doubt factor</i> for stationary scenarios only (TotDF_S)
Collaborative constructions	
CoPath_S	Correct paths inputted with the stationary scenarios: Possible range per pair=0 to 4, with four items each scoring 0 or 1
CoPath_M	Correct paths inputted with the moving scenarios: Possible range per pair=0 to 4, with four items each scoring 0 or 1
CoExpl_R	Total frequency of exchanges where one student correctly proposed that some factor was relevant and the other student accepted this proposal
CoExpl_I	Total frequency of exchanges where one student correctly proposed that some factor was irrelevant and the other student accepted this proposal

Table 2

Categories for Coding Collaborative Dialogue

Code	Label	Definition
Identify		
IF	Identify factor	A contextual parameter is identified explicitly, often at the start of a problem. E.g. 'It's the metal ball', 'There's no wind', 'It's moving'
Path		
EP	Endorse path	A suggestion is made about the path that the ball will follow, or a suggested path is accepted (including paths that are displayed onscreen). E.g. 'It's down' from one student and 'Yeah' in reply
DP	Doubt path	Doubt can range from mild uncertainty to outright rejection, but in all cases is directed at the proposed path. E.g. 'I'm gonna say no', 'I'm not sure', 'No, no, no, not that one'
LP	Link path	<i>Within</i> problems if refers to the predicted path after viewing the simulation, and <i>between</i> problems if refers to a path from an earlier problem. E.g. 'Last time I think the one on the right, but I think it's the middle'
Factor		
EF	Endorse factor	Only applies when factors are used in an <i>explanatory</i> fashion, often when justifying proposed paths. E.g. 'It wouldn't go straight because it's heavy', 'I don't think it matters if it's water'
DF	Doubt factor	Mild uncertainty to outright rejection of explanatory factors. E.g. 'Because it's lighter won't it change?' [EF] 'Weight shouldn't matter' [DF]
LF	Link factor	As with LP, can apply within or between problems, but here the link relates to factors. E.g. 'Yeah but it wasn't moving in the first one', 'So it was the same way as the metal ball'
Theory		
ET	Endorse theory	Use theoretical ideas in an explanatory fashion, both everyday notions like push, pressure, weight, and more technical notions like momentum, gravity, velocity, kinetic energy. E.g. 'Because it's got horizontal velocity probably', 'The wind the balloon is making is pushing the ball'
DT	Doubt theory	When doubt (as defined above) occurs after a remark coded ET. E.g. 'If there's a horizontal component and a vertical component, so it should be like a straight slope' [ET] 'You think so?' [DT]
LT	Link theory	When a theoretical construct used earlier (as defined for LP and LF) is referred back to
Beyond		
LB	Link beyond	Reference to something from beyond the task. E.g. 'When your plane moves forwards it looks like the thing you dropped is going backwards'

Table 3

Mean Pre- and Post-Test Scores as a Function of Age (SD in Brackets)

Age Group	Path_S		Path_M		Expl_R		Expl_I	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Year 6	3.15 _a (1.09)	3.70 _{ab} (0.47)	1.75 _a (2.27)	4.55 _b (2.40)	4.20 _a (1.85)	3.75 _a (2.31)	8.25 _a (2.81)	8.45 _a (2.61)
Year 8	2.80 _a (0.95)	3.75 _b (0.55)	1.30 _a (2.00)	2.75 _{ab} (2.27)	5.30 _a (1.69)	5.35 _a (2.01)	9.00 _a (3.43)	8.60 _a (2.56)
Year 10	2.75 _a (1.12)	3.20 _a (0.89)	1.80 _a (1.61)	1.35 _a (1.50)	4.70 _a (1.90)	4.30 _a (1.30)	8.55 _a (1.76)	8.10 _a (1.80)
Year 12	3.64 _a (0.63)	4.00 _b (0)	3.29 _a (3.65)	7.50 _c (1.09)	5.21 _a (1.67)	5.79 _a (1.19)	7.79 _a (1.72)	6.57 _{ab} (2.07)
Undergrad	3.44 _a (1.15)	4.00 _b (0)	2.56 _a (2.39)	7.56 _c (0.73)	5.81 _a (1.42)	5.50 _a (1.16)	7.69 _a (2.50)	5.62 _b (2.39)

Notes: 1) Path_S = Paths predicted from stationary carriers; Path_M = Paths predicted from moving carriers; Expl_R = Relevant explanatory factors; Expl_I = Irrelevant explanatory factors; 2) When subscripts differ within each of the 8 columns, the differences were statistically significant (Bonferroni, $p < 0.05$)

Table 4

Mean Frequency of Each Dialogue Dimension (SD in Brackets)

Age Group	Solution Confirmation	Solution/Factor Divergence	Theory Divergence
Year 6	70.20 _b (16.16)	13.40 (8.34)	3.90 (3.73)
Year 8	74.90 _b (35.55)	13.80 (7.01)	4.10 (4.80)
Year 10	40.90 _a (13.54)	6.50 (4.20)	0.60 (1.27)
Year 12	47.88 _a (11.24)	10.50 (9.26)	6.13 (4.22)
Undergrad	50.00 _a (27.27)	15.63 (13.33)	9.13 (13.88)

Note: When subscripts differ within columns, the differences were statistically significant (Bonferroni, $p < 0.05$)

Table 5

Dialogue as Predictive of Post-test Score

Path_S ($R^2 = 0.07$)

Solution confirmation	Beta = -0.12, $t = -0.87$, $p = 0.39$
Solution/factor divergence	Beta = 0.27, $t = 1.93$, $p = 0.03$
Theory divergence	Beta = 0.11, $t = 1.03$, $p = 0.15$

Path_M ($R^2 = 0.21$)

Solution confirmation	Beta = -0.41, $t = -3.27$, $p = 0.002$
Solution/factor divergence	Beta = 0.36, $t = 2.81$, $p = 0.003$
Theory divergence	Beta = 0.31, $t = 3.05$, $p = 0.001$

Expl_I ($R^2 = 0.16$)

Solution confirmation	Beta = 0.46, $t = 3.50$, $p = 0.001$
Solution/factor divergence	Beta = -0.12, $t = -0.93$, $p = 0.18$
Theory divergence	Beta = -0.20, $t = -1.88$, $p = 0.03$

Note: Path_S = Paths predicted from stationary carriers; Path_M = Paths predicted from moving carriers; Expl_I = Irrelevant explanatory factors

Table 6

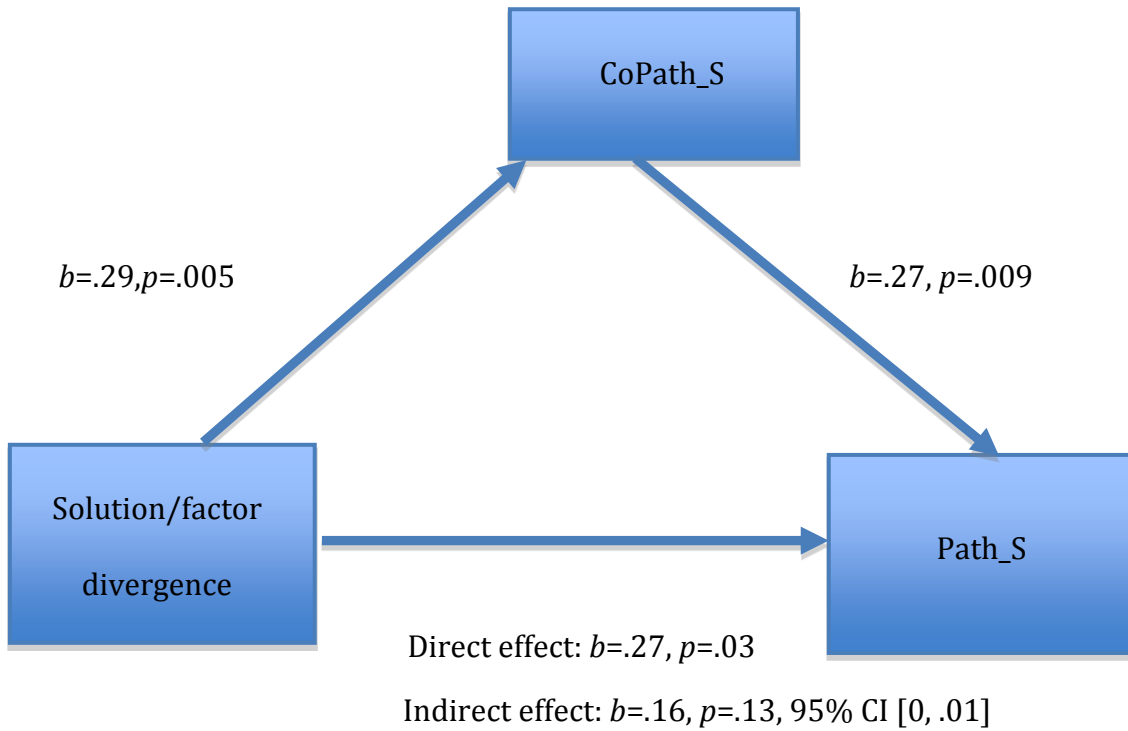
Mean Number of Accurate Collaborative Constructions as a Function of Age (SD in Brackets)

Age Group	CoPath_S	CoPath_M	CoExpl_I
Year 6	3.50 (0.71)	0.90 (1.29)	0.50 (0.97)
Year 8	3.70 (0.48)	0.60 (0.97)	0.10 (0.32)
Year 10	3.80 (0.42)	0.30 (0.95)	0.30 (0.68)
Year 12	3.63 (0.52)	0.63 (0.74)	0.25 (0.46)
Undergrad	3.88 (0.35)	1.38 (1.51)	0.25 (0.71)

Note: CoPath_S = Collaboratively constructed paths from stationary carriers; CoPath_M = Collaboratively constructed paths from moving carriers; CoExpl_I = Collaboratively constructed accounts of which explanatory factors were irrelevant

Figure 1

Relations between Solution/factor Divergence and Post-test Path_S Scores



Notes: 1) Path_S = Paths predicted from stationary carriers; CoPath_S = Collaboratively constructed paths from stationary carriers; 2) With variables non-centred, positive beta values are indicative of positive relations