

# How Technology for Comprehension Training Can Support Conversation Towards the Joint Construction of Meaning

## Abstract

Two studies assessed the role of Separate Control of Shared Space (SCoSS) technology in supporting peer collaborative discussion and comprehension. We hypothesised that providing equitable shared input to two literacy tasks (both good predictors of comprehension skill) would support discussion to promote the joint construction of meaning, and hence individual progress.

Study 1: Fifty 7- to 9-year-olds took a reading-specific multiple classification (RMC) pre-test, categorising words on two dimensions, before training on the task in pairs using SCoSS, dual-control or individual technology. Discussion produced more accurate post-test classification performance and SCoSS was associated with higher frequency of statements during training that combined both RMC dimensions (surface form and meaning of words).

Study 2: Twelve 8- to 9-year-olds were pre-tested on story recall and worked in pairs on a SCoSS-supported story construction task, requiring collaborative inference-making, hypothesis generation and selection. Post-test story recall was predicted by the frequency of deductive causal statements during training.

We discuss how technology can be used to promote collaboration and discussion that supports joint understanding and individual comprehension development.

Keywords: comprehension, conversation, multiple classification, deductive reasoning

Running head: Technology and conversation

Research into children's text comprehension problems have helped us understand much more about their underlying causes (e.g. see Paris & Stahl, 2005). Teachers may seek technological solutions to provide some form of automated support for comprehension, and considerable progress has been made in this area (e.g. the collection of chapters on automated approaches to comprehension training in McNamara, 2007).

One limitation with current work is that the technology designed to date implicitly uses individualistic models of learning and single-user technology –for example, one child working at a computer in isolation. Furthermore, in practice, because of resource limitations, two children may share a technical device that has typically been designed with a single user in mind. Such arrangements do not capitalise on the strengths of sharing expertise and may result in unhelpful domination and inequitable contributions of users. Individualistic models do not sit well with a tradition of training studies using paired reading, group work and discussion. For example, Palincsar and colleagues pioneered the use of reciprocal teaching (e.g. Palincsar & Brown, 1984), and a fundamental part of this is the conversations children have with each other about the text.

Reciprocal teaching was originally underpinned by Vygotskian developmental theory, which stresses the fundamental role of social interaction in cognitive development. Higher-order functions (in this case, comprehension) arise from language-based social interaction. An important function of a teacher or more capable peer is to support a student of lower comprehension ability so as to reach a higher level of competence than could be achieved alone (the zone of proximal development). Gradually, the less able partner takes over the more able partner's role and becomes able to perform tasks independently.

The role of discussion is central to this developmental process, but has not been extensively studied in the area of reading comprehension. In the Vygotskian framework, sharing knowledge and explaining one's own understanding is of positive benefit to both members of a pair. Both producing and hearing explanations help learners to move from performing cognitive functions between people (interpsychological) to the individual level

(intrapyschological). Siegler (2005) has argued that self-explanation – seeking an understanding of why one answer is correct and others are incorrect – improves learning in a wide range of domains, such as maths and science. Studying such explanations that arise in peer discussion during comprehension training also brings benefits to the researcher: it provides us with a window into the developing processes of children’s thinking, and in particular, their comprehension. One example of this approach is a study by Yuill (in press), showing that changes in children’s comprehension scores after discussion-based training was predicted by increases over sessions in children’s use of metalinguistic terminology. This suggests that contextual support for children to explain their understanding to a peer may help children to develop, through discussion, a reflective attitude to language. Given that there is typically large variation in children’s responses to training, studying the content of conversation as a predictor of children’s improvement may provide useful clues about processes of comprehension change.

Discussion can operate in one of two ways, as described by Warschauer (1997): as the teacher modelling an approach to learning, or as a way of generating new meanings collaboratively (e.g. Wertsch & Bivens, 1992, cited in Warschauer, 1997). In the current paper, we focus on support for collaborative discussion, drawing on recent work in human-centred technology (HCT), in order to develop and evaluate software that promotes collaboration as a means to support individual cognitive change in comprehension processes.

The concept of collaboration is a vital part of the design of software that is informed by the Vygotskian approach taken here. In the HCT literature, Teasley and Roschelle (1993) drew an important distinction between collaboration and cooperation. Cooperation involves a division of labour, with each person independently responsible for one part of a task, whereas collaboration requires partners to coordinate and maintain a shared understanding of the task. This push towards shared understanding (intersubjectivity) is a crucial aspect of collaborative learning that ties it to Vygotskian theory. Kerawalla et al. (2008) argue that collaborative learning requires greater explanation of one’s own thought processes, thus encouraging self-understanding, self-regulation and learning through conversation. Such learning can promote

better shared understanding of a problem space than cooperative learning, where tasks are completed in parallel, and effort is required to manage turn-taking. Pearce et al. (2005) introduced a general theoretical framework, the task-sharing framework (TSF) to describe how collaborative tasks might be shared between users. This paper describes the possible benefits of two pieces of software that use the TSF in different ways to support collaborative working in reading comprehension tasks. In Study 1 we use an interface that allows users both to have control of a shared task space, and in Study 2 we use software that allows two users collaboratively to explore a space of possible interpretations of ambiguous text, using the principle of state expansion (the PoSE approach: Pearce & Luckin, 2007.) The particular comprehension tasks used here were chosen because they were judged to be suitable material for the particular method of task-sharing, they correlate highly with comprehension skills, and had been used as training techniques for improving comprehension. Children in the age range of 7 to 9 were chosen because of the large body of previous research on different comprehension tasks in this age range (e.g. Yuill & Oakhill, 1992).

The interface developed from the TSF was designed to encourage shared understanding while still representing and allowing individual control over each individual's conception of a task. This interface was designated Separate Control of Shared Space (SCoSS: Kerawalla et al, 2008). SCoSS has four key features to support collaboration and shared understanding, shown for a sample task in Figure 1. First, the users have a single task to solve and on the screen they can see their own task state and their partner's task state. Second, they have their own mouse, with independent control of their own task space, and no control over the other's space, thus discouraging the type of domination often seen when sharing single-user technology (e.g. grabbing the mouse and changing the other's work). Third, the extent of agreement and disagreement is shown explicitly: for example, agreement is represented by green shading of the agreed area in each user's task space. With these features, it would still be possible for users to complete the tasks independently, ignoring the agreement indicators. The fourth feature of SCoSS reduces this, by having points in the task

at which users have to agree in order to move on, or to get the next piece in the task. This need for agreement pushes for discussion, in particular about areas of disagreement.

Figure 1 about here

In this paper we assess the role of SCoSS technology in supporting two different literacy activities, the outcome of training on the individual and the relation between children's collaborative discussion during training and the outcome in two quite different literacy tasks which both correlate highly with reading comprehension skill: in Study 1, reading-specific multiple classification tasks ( Cartwright, 2002), and in Study 2, a clue-word inference training task (Yuill & Joscelyne, 1988).

#### Study 1: Using SCoSS to Facilitate Discussion about Semantic and Grapho-phonemic Aspects of Words

Our first example of using SCoSS to support literacy draws on the idea that good reading comprehension is underpinned by cognitive flexibility. Cartwright (2002; 2008) has argued that reading with understanding requires the flexible coordination of grapho-phonemic and semantic processes: to be good comprehenders, children need to coordinate both sets of processes. Cartwright presents evidence that the ability to coordinate form and meaning is a strong predictor of children's reading comprehension skill. The acquisition of such cognitive flexibility in children has been studied both as part of a naturally occurring developmental process (Piaget & Inhelder, 1966/1969; Sternberg & Berg, 1992) and as a skill which can be actively developed through instruction and practice (Bigler & Liben, 1992).

Cartwright's work demonstrates the role of cognitive flexibility in reading using standard measures of multiple classification: classifying objects or words on two dimensions simultaneously. More specifically, she has shown that reading comprehension measures are better predicted by multiple classification performance that involves reading, rather than general classification tasks (Cartwright, 2002). The reading multiple classification task (RMC) requires children to sort words on 2 dimensions: grapho-phonemic and semantics. For example, the 12 words in Figure 2 are to be sorted into a 2x2 table with 'ch' vs 't' words in the columns and 'body parts' and 'food' in the rows. This is a classic Piagetian task, and

children do not find it easy: correlations with reading comprehension can be demonstrated throughout ages 5 to 11. Cartwright sees this classification skill as part of a more general executive function skill of cognitive flexibility, prefigured, for example by the difficulty that preschoolers have in executive function skills such as classifying objects successively by colour and type, in the dimensional change card-sorting task (e.g. Kloo & Perner, 2005).

Figure 2 about here

Such tasks are generally thought of as measuring individual, stable skills, but as Cartwright (2002) has shown, RMC skill can also be improved through practice: training individual children using repeated practice of RMC tasks leads to increases in RMC skill and some transfer to reading comprehension skill. Based on our reasoning above, we would expect that collaborative discussion of RMC tasks, using the SCoSS framework to support such discussion, could also produce improvements in RMC skill. If discussion promotes internal cognitive change, then the type of discussion children have during the task should be related to the rate of change in RMC skill. Study 1 is an initial exploration of the potential of such technology to support discussion and cognitive change. As the training in this initial study was very limited, we did not expect large training effects, but also wanted to investigate differences in the type of discussion children have. Cartwright used scores reflecting both speed and accuracy: given that our discussion training involved deliberation with no emphasis on speed, our main performance measure was children's accuracy independently of speed, rather than Cartwright's traditional accuracy/speed measure.

## Method

*Participants.* Sixty-four 7- to 9-year-old children (32 girls, 32 boys) at a primary school in a small town in south-east England were recruited through informed consent from parents and children. Children were screened to ensure adequate reading ability for the test materials using the Gates-MacGinitie Vocabulary Test Level 2 Form K (1989) as a measure of word reading. For the training session, children were paired with a same-sex peer, using pre-test

RMC scores, so as to have mismatched pairs (one higher RMC score, one lower), in line with the Vygotskian idea of having a more able peer to support the less able one. We also checked with the teacher that no pairs comprised children who, in the teachers' opinions, would not work together well. Pairs were randomly allocated to a SCoSS or a dual-mouse, single-user interface condition, as described below. Our analyses required complete data for a series of 5 separate sessions for both children in each pair, and because of absence, unavailability or technical failure we had complete data for only 40 children. We could detect no difference between the complete and reduced samples on any variable (age, vocabulary, gender, pre-test or post-test speed or accuracy scores) and given the many different reasons for data unavailability, it seems unlikely this represents a single source of significant bias. It also allows us to complete all analyses on the same dataset.

A further 10 children acted as no-discussion controls, completing the pre- and post-tests and the training items individually with single-user software. They were recruited from a different class so that they did not experience the discussion training. Children were not distributed into classes by the school according to criteria that might be expected to affect the results, e.g. ability.

We therefore had 3 groups of children, whose characteristics are shown in Table 1. The three groups did not differ significantly in vocabulary score or RMC pre-test scores (either accuracy or accuracy/speed, see below), all  $F_s < 1$ . There was a marginal though not significant effect of age between the 3 training groups,  $F(2, 49) = 2.72$ ,  $p < .07$ , because the no-discussion controls were slightly younger, so as a precaution age was included as a covariate as appropriate in the analyses.

Table 1 about here

*Procedure.* Trained children had 3 separate sessions: a group-administered individual computerised RMC pre-test, followed by a training session in a pair, which was videotaped for further analysis, and then a group-administered individual RMC post-test. Pre- and post-sessions were administered within a week of the training session. Before the training, the two

researchers had worked with the two teachers and classes involved in the paired conditions, with two lessons delivered focusing on principles of equitable discussion (see Dawes, Mercer & Wegerif, 1990). Children in the single condition had 3 separate RMC sessions, as above, but were not videotaped.

*Pre- and post-tests of RMC.* In each session, there were three sets of 12 words to be classified and a further practice set for the first session, all presented on laptop computers. Children were given instructions adapted from Cartwright (2002), wherein the instructor demonstrated a correct sort and explanation, noting that the words were to be sorted ‘two ways at the same time: one way is to do with the word’s sound and the other way is to do with meaning’. In this phase, the grapho-phonemic rule was always the first sound in the word, but the semantic dimensions varied (e.g. including food, clothing, containers). As in Cartwright’s studies, some dimensions were defined by the absence of an attribute (e.g. containers vs. not containers). The first screen presented all 12 words and a 2 x 2 grid, and children had to drag words into the grid, with no cell on the grid accepting more than 3 words. Words could be moved between cells on the grid after initial placement, if needed. On completion, children recorded the reason for their classifications on a Dictaphone, and these were later scored by 2 independent raters with high agreement (over 95%). The software recorded correctness of classification and time taken to complete it.

*Training.* We used a specially-developed piece of software, WordCat, to present the RMC task. Following a guided practice round of 12 words to familiarise the children with the software, there were 3 further sets of 12 words. Training sets were deliberately made more difficult than the pre-test in three ways. First, we used four different grapho-phonemic rules, e.g. number of letters, presence vs. absence of double letters, rather than initial letter. We warned children the rules would be different from the pre-test. Second, children were given one word at a time to place rather than seeing the whole set initially. Third, we ensured that the words appeared in an order that could suggest ‘red herring’ classifications, so that incorrect sorting inevitably occurred and words had to be re-categorised. Pilot testing showed these to be effective means of generating discussion and justifications for decisions. Children

therefore had to think hard about what the semantic rules might be, and what surface features differentiated the words. As Cartwright specified, the task requires the concurrent use of grapho-phonemic and semantic properties of language. After each child had placed the current word, a screen message asked them to discuss and agree on the placement.

*SCoSS condition.* Each child had their own task representation and set of words, on opposite sides of the screen. Children could only move words in their own representation and could not get the next word until they had agreed on placement of the current word: both had to click their own 'agree' button.

*Dual-control condition.* This involved two mice controlling all areas of a single grid, so that each child had their own mouse which could control everything on the screen (e.g. moving words, clicking 'agree' buttons). Children could therefore over-ride each other's decisions.

*Individual condition.* One child had a single task representation and one mouse which could control all the actions required to complete the task.

At the end of each 12-word task, there was feedback on the computer screen, with the words being corrected if sorting was incorrect, and an explanation of why each subset of words went together. Paired training sessions took 30-40 minutes and were video-recorded for further analysis.

*Scoring.*

*RMC task.*

*Accuracy score:* Following Cartwright (2002), explanations of sorting gained an accuracy score of 0-3 for each of 4 rounds in the pre-test and in the post-test, giving a maximum of 12 points for each: 3 = correct sorting and explanation, mentioning both grapho-phonemic and semantic dimensions, 2 = incorrect sorting and correct explanation given after the words had been automatically corrected by the software, 1 = correct sorting but incorrect explanation, 0 = incorrect sorting and incorrect or no explanation.

*Accuracy/speed score:* Following Cartwright (2002), accuracy scores were divided by the sorting speed (in seconds) recorded by the software, and the result multiplied by 100 for ease of working.

*Explanations during training.* We calculated the number of explanations of each type for each trained pair. We coded explanation types by pair rather than by individual, because if SCoSS works as intended, children are more likely to produce joint explanations than in the dual condition (see Kerawalla et al., 2007, for examples), therefore it is inherently more difficult to allocate an explanation as originating from an individual child in the SCoSS than in the dual control condition. Each explanation offered during the training session was categorised in two independent ways, as follows:

*Surface form or meaning:* reference to the surface form of the word, i.e. letters, sounds or spelling of words e.g. ‘they all begin with ch’, or to the meaning, e.g. ‘they’re all kinds of animal’.

*Uni-polar or bi-polar:* reference to just one end of the dimension, e.g. ‘they are animals’ or both ends: ‘Those begin with ‘b’ and those begin with ‘s’’.

Occasionally, children also gave *combined* explanations, mentioning grapho-phonemic and semantic properties in a single utterance, e.g. ‘the words in this box have four letters and they are types of weather’. Because these were comparatively rare ( $n = 33$ ), we did not include them in the main analysis but we report an overall analysis of totals of this type.

## Results

### *RMC Performance.*

*Accuracy.* An analysis of variance of the RMC accuracy scores at post-test with group (SCoSS, dual-control or individual) between subjects and pre-test accuracy score and age as covariates showed a main effect of group,  $F(2, 44) = 5.23$ ,  $p < .01$ ,  $\eta^2 = .18$ . Simple contrasts showed that the (age-corrected) mean for SCoSS post-test accuracy (7.78, s.d. 0.6) was significantly higher ( $p < .005$ ) than the individual condition (corrected mean 3.7, s.d. 1.1) as was dual-control accuracy (corrected mean 7.31, s.d. 0.8), although the latter was not significantly different from the SCoSS mean.

*Accuracy/speed:* A similar analysis of variance of converted difference scores of speed and accuracy comparing training condition (SCoSS, dual or individual), with age in months as a covariate, showed no main effect of condition,  $F < 1$ .

*Explanations.* As expected, improvement was variable. This is not surprising given that children were relatively free to interact with the material as they wished. This variability, and the high numbers of explanations generated (mean = 23.3 per pair) allows us to analyse, for the SCoSS and dual-mouse conditions, what sorts of conversations seem most useful in helping children coordinate the different properties of the words.

Total number of explanations of all types was significantly higher for SCoSS pairs than for dual-mouse pairs, respective means = 46.6 (s.d. 17.8) and 22.5 (s.d. 13.0)  $F(1, 38) = 21.45, p < .001, \eta^2 = 0.34$ , and the correlation of number of explanations with the extent of improvement was moderately positive,  $r(33) = .34, p < .06$ . In particular, pairs in the SCoSS condition gave significantly more bi-polar explanations than dual-mouse pairs, means = 16.42 (s.d. 17.6) and 4.38 (s.d. 3.4) respectively,  $F(1, 38) = 7.25, p < .01, \eta^2 = 0.15$ . Finally, ‘combined’ explanations, although relatively rare, occurred on 29 occasions in the SCoSS condition compared to only 4 in the dual-user condition.

### Discussion

As predicted, paired conditions with encouragement to discuss classifications produced greater improvement in accuracy of individual word classification performance on a post-test, compared to individual practice with no discussion. There were differences between the SCoSS and dual-control pairs in the *type* of explanation. SCoSS pairs more often used explanations that combined both dimensions in a single utterance, and were more likely to use classifications that mentioned both ends of a dimension, although this did not yield significantly better performance on individual accuracy scores at post-test. The ability to compare directly any differences between one’s own and one’s partner’s classification, and the need to agree, may have encouraged children in the SCoSS condition to be very explicit about the reasons for classification.

Although the training here was very brief, serving as a research tool rather than a suggested educational practice, it seems that SCoSS-supported discussion could engage children in discussions that might be used to improve RMC performance, and possibly cognitive flexibility generally. It may be that these explanations themselves are underpinned by individual cognitive skills, for example working memory: it was clearly quite challenging for children in the current study to hold in mind the different dimensions and to construct an utterance that mentioned both. Working memory is clearly an intrapsychological variable, but anecdotal evidence from our study suggests that this too may benefit from interpsychological processes. We noticed that occasionally, children using SCoSS would share responsibility for the different dimensions; so for example, one child would mention one dimension and the other child would give the other, sequentially or very close together. Given the comparative rarity of ‘combined’ explanations mentioning form and meaning in the same utterance, it seems that collaborative ways of working might help children to share responsibility for a complex task which later needs to be completed in a single child’s head. The following example discussion, from a pair using SCoSS, shows how this might work, and illustrates a collaborative interaction, as opposed to a cooperative division of labour. Callum consistently takes responsibility for one dimension and Fred for the other, although the children alternate in their mention of form or meaning, clearly able to consider either one or the other, but having trouble with dealing with both simultaneously.

*Callum: Finger should go in there because it’s another part of the body [meaning]*

*Fred: And it begins with f [form]*

...

*Callum: Lettuce would go under lolly because it’s l. [form]*

*Fred: Food! [meaning]*

*Callum: And it’s food*

...

*Callum: Lung would go in there because it’s part of the body.[meaning]*

*Fred: And begins with l [form]*

Towards the end of the session, Fred suddenly is able to describe both form and meaning, which he does emphatically:

*Fred: Chick would go in there because it begins with ch AND it’s a type of bird*

*Callum: Yeah...*

*Fred: Yes, AND! [highlighting the need to mention two dimensions]*

The fact that SCoSS engendered more ‘combined’ explanations suggests that it does as intended encourage children to construct joint meaning in a shared task. A more microgenetic analysis of repeated sessions might perhaps show the extent to which such combined working would eventually lead to an individual child being able to describe both dimensions in a single utterance alone, rather than requiring scaffolding from a partner.

Kerawalla et al. (2008) provided a qualitative analysis showing how children using SCoSS are enabled to express their own ideas, but also need to take account of a partner’s views, and either have to justify their own views or adopt the partner’s view, along with the partner’s justification. The quantitative data here show clearly that SCoSS engenders a greater number of complex justifications, compared to dual control, and improved accuracy on the task compared to individual practice with feedback. Our study was clearly limited, and further work is needed to assess more extended training and whether any effects would extend to traditional reading comprehension tasks.

We have not statistically analysed the possible role of peer domination in SCoSS vs. dual control, but the nature of the different interface designs simply precluded certain behaviours. For example, children could work independently on different parts of the task in the dual-control condition, whereas children with SCoSS both had to agree before they could move on. Domination or bullying another child to agree was almost unseen: it would have required one child moving over and taking the other child’s mouse, and this is extreme dominance that we saw in only one instance. Domination was afforded more easily in the dual-control condition because each mouse could perform the same actions anywhere on the screen, therefore either child alone could repeatedly undo the other’s work, or place a word where they wanted it, without discussion, and move on to the next word without reaching true agreement.

One possibility is that SCoSS was specifically helpful in the word categorisation task because of the support for combining explanations: the task lends itself to sharing of responsibility. As we might expect, training may have helped children to understand the

general conceptual aspect of the task (2 by 2 classification), but discussion did not produce improvement in the usual RMC measure combining speed and accuracy: deliberation clearly did not help speed. We now turn to a second, more open-ended implementation of SCoSS in a more reflective task of understanding the relations between inferential conclusions from textual evidence, , to see whether fostering collaborative discussion would support comprehension skills more directly.

### Study 2: The Role of SCoSS in Understanding the Relation Between Evidence and Conclusions

The word classification task in Study 1 taps a general cognitive capacity that is thought to underlie good reading comprehension. In the second study, we investigate the use of SCoSS in a training task that directly involves comprehension, awareness and monitoring skills. One skill that children with poor comprehension seem to lack is that of interrogating a text, even when they have the prerequisite cognitive requirements, such as knowledge and memory. For example, Cain & Oakhill (1999) showed that poor comprehenders sometimes failed to answer inference questions even when they were shown the relevant sentence in the text and when they demonstrably had the knowledge required to make the specific inference.

One training technique that has been used to support inference-making in poor comprehenders is the 'clue word' procedure developed by Yuill & Joscelyne (1988). Poor comprehenders of 7- to 9 years of age were given highly ambiguous texts and were trained to search for words that would give them clues to interpreting particular aspects of the story, such as the setting and main event. Children given such training improved significantly more than non-trained in the recall and comprehension of similarly ambiguous stories in a post-test. The original study involved individual children working out the clues, guided by an adult. For example, for the sentence 'Billy was crying: all his work had been ruined by the wave', the adult guided the child to think what the 'wave' might be, and how it might be related to work being 'ruined', what the location of the story might be, and so on.

This task requires exploring the space of possibilities in interpreting the ambiguous text, a potentially difficult task, as addressed by Pearce and Luckin (2005) in the PoSE framework. A crucial way of addressing possible interpretations is to look at how the evidence might constrain interpretation: some possible interpretations can be ruled out. For example, mention of the word 'sand' would help constrain the interpretation of 'wave' in the example above. We thought that software using SCoSS technology could be used to help pairs of children co-construct the meaning of a puzzling story through selecting and appropriately rejecting different interpretations using the textual evidence. Just as the conversation between child and adult picked out clues and used deductive reasoning, so children might be able to support each other in such reasoning, given appropriate technical support.

Puzzling stories seem to be helpful in training because their obscurity is apparent to children, who realise that they must work to interpret the meaning. Our second implementation of SCoSS technology, WordBird, was designed to provide children with pictorial representations of puzzling stories, graphically illustrating different possible interpretations of an unclear text, so that children could work out which interpretation was most consistent. If both children in a pair constructed their own possible interpretations, SCoSS technology would allow them to see the other's interpretation, to compare it with their own and to discuss which interpretation might be more consistent with the text, so as to reach joint understanding. This method fitted well with the puzzle stories originally used by Yuill and Joscelyne, which were written so as to provide multiple garden-path interpretations that could only be resolved as further sentences were revealed to the children. Multiple possibilities had to be kept in mind, and rejected as evidence against them or other unconsidered interpretations became apparent. The software was designed to assist in this process, using the PoSE framework. As the software was more complex to use than WordCat, we used children in the older part of our specified age range.

This type of task presents reading as a deductive exercise, used in programmes such as the reading detective approach (e.g. Paris & Jacobs, 1984). The children's task is to search

for evidence and to use it to support or reject particular conclusions. Doing this task in a SCoSS-supported pair requires children to compare and justify interpretations. Such justifications need to draw heavily on deductive reasoning, e.g. 'I think the story is set by the sea because 'wave' could mean sea water'. We know that children with poor comprehension are weak at deductive reasoning of just this sort. Oakhill, Yuill and Donaldson (1990) assessed comprehension and production of deductive 'because' in comparison to empirical use of 'because' in short text scenarios, e.g. 'I know Mary has a cold because she is sneezing' (deductive) vs. 'Mary has a cold because she stayed outside in the rain' (empirical). Children with poor comprehension did as well as good comprehenders in producing and understanding empirical uses of 'because', but they were significantly poorer in deductive uses (questions starting, 'How do you know that  $x$  is the case?'). However, a further condition in the Oakhill et al. study showed that poor comprehenders could improve given support. They performed significantly better in a new condition where a picture was used to draw additional attention to the evidence supporting a deductive inference. For this reason, we felt that supporting collaborative discussion between peers about deductive reasoning using pictures would be an appropriate use of SCoSS technology, if it supports comparison and justification of knowledge in a non-dominating and engaging environment and provides a need to agree on a joint solution reached through collaboration.

We applied SCoSS to the clue-word task, by developing the WordBird software, using the task-sharing framework (Pearce et al., 2005) to give two users a shared space to place pictures representing possible interpretations of an unclear text. Each child could see but not alter the other's pictures, and the owner of each picture was identified by the colour of the picture's border. Text was presented sequentially, so that the pictures on the desktop representing the story could be amended, kept or discarded as new textual evidence appeared. For example, one story mentioned a 'ride'. When children clicked on this word, they were given pictures of a horse or a fairground big wheel. A subsequent sentence refers to a 'wheel', so the picture of the horse may then be discarded and the fairground ride kept. Unlike WordCat, there was no indication of agreement (other than comparing each others' pictures)

and no need to agree at specified points. However, the final aim of the task was for children to agree on an identical single representation that summarised the situation in the story. A small progress indicator at the foot of the screen showed children how close they were to agreement, but a lack of agreement did not prevent them from continuing to explore the space of possibilities. We expected that the task would support discussion of what inferences could be drawn from the text, and that as in the study by Oakhill et al., a focus on information in pictorial form would support deductive inference. We therefore expected increases in story comprehension from pre- to post-test, and that improvement should be associated with greater use of deductive statements. We were also interested in studying more qualitatively how children used deductive statements, and whether poor comprehenders would be able to do so given support.

#### Method

*Participants.* 16 children (10 boys, 6 girls) from Year 4 (age 8-9) of two schools in a city in south-east England were recruited through parental and child consent. Children were selected from a wider group on the basis of their comprehension scores (see below) so that we could pair children differing in comprehension skill. The characteristics of the children are shown on the left side of Table 2. The two groups were not significantly different in chronological or reading accuracy age, but differed significantly in comprehension age.

Table 2 about here

#### *Materials*

*Reading assessment.* Children's reading was assessed with the Neale Analysis of Reading Ability (Neale, 1989), to yield a reading accuracy score (based on decoding errors for reading passages of narrative text) and comprehension (based on answering questions on the text that had been read). Using the criteria of Oakhill et al. (1989), children were judged to be poor comprehenders if their comprehension age was at least 6 months below their accuracy age, and not more than two months above their chronological age.

*Pre- and post-tests.* Six puzzling stories, modelled on those used by Yuill & Joscelyne, were created, and randomly divided into 2 sets of 3 for pre- and post-test. Each story consisted of 10 idea units and was followed by 5 questions requiring inferences.

*Training.* Four puzzling stories based on those used by Yuill & Joscelyne were created in such a way that pictures could gradually be built up to represent the situation. We deliberately inserted ambiguity and false trails into the text so that children had to backtrack and revise their interpretation as sentences in the story were successively revealed in a text box at the top of the screen. The children's task was to build up a picture in their workspace that represented the final situation described in the story, for example, a boy on a big wheel dropping his ice cream onto a passer-by's hair. Both children shared the lower part of the screen as a workspace in which to build up pictures from pre-constructed elements (e.g. picture of a boy on a big wheel, of a girl with messy hair) and each child's constructed pictures were identified with coloured borders. Children could have as many or as few pictures as they liked in the shared space, could add elements to pictures and remove them, and add new pictures and remove their old pictures from the workspace. A box at the top of the screen contained the text, revealed a sentence at a time. In each sentence, certain 'clue' words were highlighted. When these were clicked, pictures representing possible interpretations were made available, making a clear link between an ambiguous piece of text and different possible pictorial representations of that text. For example, given a starting picture of a boy (the main character), clicking on the next clue word 'ride' added two pictures building on the existing picture of the boy, showing one picture with the same boy riding a horse and another with the boy riding on a big wheel.

*Procedure.* Children were tested individually in a quiet room for the pre- and post-tests within ten days either side of their two training sessions, which took place within one week of each other. For the pre- and post-test, the tester read aloud each story from a booklet, and immediately afterwards asked the child to recall as much of the story as possible, and then to answer comprehension questions. The two training sessions were conducted with the children in pairs in a quiet room with one of two trainers, who had completed brief training together.

This training ensured that the trainers explained the software in the same way, checked children could use it competently, and provided the same level of support if children had difficulties. Children completed two puzzling stories in each training session. Sessions were videotaped and typically lasted 30 minutes for session 1 and 15 minutes for session 2.

### *Scoring*

*Pre and post testing.* Recall scores were calculated as the total number of idea units recalled over all 3 stories, a maximum score being 30. Comprehension questions yielded a score of 5 for each of the 3 stories, with a total maximum score of 15.

*Explanations.* Each session was transcribed and coded for the number of times each child used deductive 'because' statements, which occurred primarily in justifying actions such as placing, adding to or removing a picture from the workspace, for example, 'It can't be a beach ball because it's not hard enough (to break a window)'. Occasionally the word 'because' was only implicit, if children were responding to a 'why' question.

## Results

### *Recall and Question-Answering*

There was a significant increase in question-answering score from pre-to post-test,  $t(14) = 2.5, p < .05$ , but not in recall scores,  $t(14) < .28, n.s.$  An analysis of variance of the post-test scores between groups, with pre-test score as a covariate, showed no effect of group for post-test story recall,  $F < 1$ , which had not differed at pre-test either. The pre-test difference between groups in question-answering was no longer apparent in the post-test,  $F < 1$ , with adjusted means for post-test questions of 12.39 for poor and 12.13 for good comprehenders.

### *Explanations*

A mean of 11.43 (s.d. 5.2) uses of deductive 'because' per child occurred over the two training sessions. Session 2 was generally about half the length of Session 1, mainly because of increased familiarity with the software and improved grasp of the task, with fewer false leads and unnecessary moves being taken. In keeping with this ratio, the number of uses of deductive 'because' was also approximately twice as high in Session 1 as in Session 2.

With appropriate caution for the volatility of correlational data in such a small sample size, we then looked at whether deductive statements in conversations during the training sessions would predict changes in story recall and question-answering. The number of deductive statements used over the 2 sessions correlated with changes in story recall,  $r(14) = .81$ ,  $p < .001$ , but not with changes in question-answering,  $r(14) = .32$ , n.s. Use of deductive 'because' did not correlate with children's age, accuracy or comprehension scores, all  $r_s(14) < .2$ .

### Discussion

Previous research showed that the 'clue word' approach can be an effective technique for increasing comprehension. The present study is a modest pilot exploring how technology might support discussion of text interpretations by supporting the construction of alternate meanings, and allowing comparison of one's own interpretation with another child's. The need to reach agreement may have encouraged children to explain their own reasoning, and to listen to the other's reasoning. Particularly notable was the high level of use of deductive reasoning, linking the form of the text to the interpretation of meaning. Frequency of this predicted increases in later story recall, although it did not predict increases in question answering.

The results complement the findings of Cain (2007), who found that children who had to justify their answers to an adult did better in a word definition task than children who just had feedback on their answers. Cain suggests that explanation may have directed children's attention to the text. The current study made the connection between text and inference very clear, by providing (unambiguous) pictorial information linked directly to specific ambiguous words and phrases in the text. Children's conversations suggested that this link prompted them to use deductive reasoning freely in choosing or rejecting pictures. For example, Dan, a poor comprehender, was asked why he was deleting a picture of a beach ball by a window: 'cos beach balls are soft... it could be a tennis ball or a football because they're hard' [hard enough to break a window]. Another strategy that was used seems more sophisticated than this, in that children sometimes referred explicitly to the text as a warrant

to draw or reject particular inferences. For example, Liam, a good comprehender, rejected a picture of a boy standing by a sun lounger ‘because it says he’s lying down’, and Amy, another good comprehender, explained ‘because it says he took the ice cream on the ride...if you look {pointing at the text} it will tell you {what you need to know}’. This suggests that the task encouraged children to take the reflective metalinguistic attitude that Yuill (in press) has argued that such children lack. However, such statements do not guarantee a correct inference: for example, Simon (good comprehender) rejected the ‘horse’ picture for the ‘ride’ clue because ‘it doesn’t say *horse*’, so it is still possible to hold misconceptions despite a reflective attitude.

Although this study did not include a comparison non-SCoSS condition, the discussions may have been helped by the SCoSS technology we used, which enabled children to have control over their own work space and the facility to compare their solutions with those of a partner. Although the task was a challenging one individually, children sometimes used their partners’ pictures and reasoning to amend their own. It is interesting to note that even from the first session, both good and poor comprehenders were using deductive statements. Just as Oakhill et al. (1990) found that drawing attention to pictorial evidence led to improved understanding of deduction in poor comprehenders, so the current study showed that poor comprehenders can produce deductive reasoning, sometimes quite complex, given appropriate support from the task, and can discuss this reasoning with their peers. This is the sort of conversation that could be part of strategies such as reciprocal teaching, and technology such as SCoSS might foster equitable discussions of this type.

#### General discussion

In two studies, we examined the role of sharing knowledge and working collaboratively on children’s ability to perform reading-related tasks. In Study 1, we found that SCoSS technology requiring frequent agreement through discussion between children led to improved RMC performance, in comparison to individual training but not to a dual-control condition. The SCoSS condition did, though, produce more complex justifications, combining grapho-phonological and semantic properties, than the dual-control technology condition. We

speculated that SCoSS provided conditions that afforded equitable discussion and the need to justify word classifications, in a way that might be used to support comprehension skills.

Given that this was a small-scale study, replication of these findings is needed, and it is important to see whether SCoSS would have advantages on different outcomes, such as tasks directly tapping comprehension skills, and follow-up over a longer period of time.

Study 2 assessed the use of SCoSS technology adapted to a task requiring the discussion of different interpretations of highly ambiguous text, in which interpretations had to be justified and constraints applied in order to reduce the space of possible interpretations. Children who used more deductive statements during the training discussions tended to be those who did better on individual post-test story recall. Given the difficulties that poor comprehenders have with use of deductive statements to derive conclusions from textual evidence, we found a high rate of uses of deductive inference. We speculated that the use of SCoSS technology, which focuses attention on shared understanding but allows independent decision-making, provided a stimulus for children to explain their reasoning to each other. Existing literature, and Siegler's work in particular shows that self-explanation can be a powerful engine of development.

We tend to think about comprehension as an individual cognitive process, but these studies suggest that conversation could be an important means through which children can negotiate and justify their interpretation of text. This is, after all, what can drive the interpretation of spoken text, where there are many subtle cues that interlocutors can use, both to signal and to clarify misunderstanding, such as looks of puzzlement, gesture, subtle alterations in style of speech and eye contact. These cues are not available in text, where the author is communicating with the anonymous reader, and interaction is relatively one-way. While we may ask children to 'interrogate' a text, the text itself is hardly able to answer back, except in the most passive way. A peer, on the other hand, who is struggling to understand the same text, is much more of an 'open book' as s/he shows uncertainties and hesitations, false trails and realisations. In these studies, we sometimes observed definite 'aha' experiences, where children suddenly came to an understanding through discussion: these experiences

serve a purpose for the speaker and for the listener, as well as for the observing researcher, who has an external representation through dialogue of children's thought processes while reading and interpreting text.

Such a view of text is consistent with what Warschauer (1997), following Wertsch and Bivens (1992), describes as the text-mediational view, with 'texts as 'thinking devices' to generate new meanings collaboratively' (Warschauer, 1997, p.471). Warschauer cites the work of Wells and Chang-Wells (1992), who urge that texts are treated as 'a tentative and provisional attempt on the part of the writer to capture his or her current understanding...so that it may provoke further attempts at understanding as the writer or reader dialogues with the text in order to interpret its meaning' (ibid., p. 471). Further work could illuminate the role of conversation in text comprehension, and of how technology can support this conversation.

## References

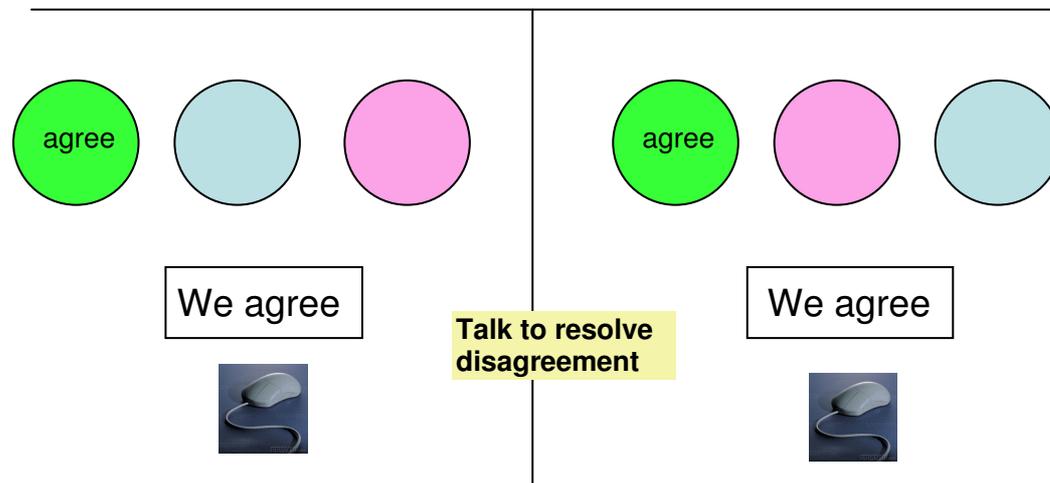
- Bigler, R. S., & Liben, L. S. (1992). Cognitive mechanisms in children's gender stereotyping: Theoretical and educational implications of a cognitive-based intervention. *Child Development*, 63, 1351-1363.
- Cain, K. (2007). Deriving word meanings from context: does explanation facilitate contextual analysis? *Journal of Research in Reading*, 30, 347-359.
- Cain, K. & Oakhill, J. V. (1999). Inference making and its relation to comprehension failure. *Reading and Writing*, 11, 489-503.
- Cain K.; Oakhill J. V.; Barnes M. A.; Bryant P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, 29 (6), 850-859
- Cartwright, K. B. (2002). Cognitive development and reading: The relationship of reading-specific multiple classification skill to reading comprehension in elementary school children. *Journal of Education Psychology*, 94, 56-63.
- Cartwright, K.E. (2008). (Ed.) *Literacy Processes: Cognitive Flexibility in Learning and Teaching*. New York: Guilford.
- Dawes, L. Mercer, N. and Wegerif, R.(1990) *Thinking Together*, UK: Imaginative Minds.
- Gates, A.I. & MacGinitie, W.H. (1965). *Gates-MacGinitie Reading Tests*. New York: Columbia University Teachers' College Press.
- MacGinitie, W.H. & MacGinitie, R.K. (1989). *Gates-MacGinitie Reading Tests*. Chicago: Riverside.
- Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child* (E. A. Lunzer & D. Papert, Trans.). New York: Humanities Press.
- Kerawalla, L., Pearce, D., Yuill, N., Luckin, R. & Harris, A. (2008). "I'm keeping those there, are you?" The role of a new user interface paradigm – Separate Control of Shared Space (SCOSS) – in the collaborative decision-making process. *Computers and Education*, 50, 193-206.

- Kloo, D. & Perner, J. (2005) Disentangling dimensions in the dimensional change card-sorting task. *Developmental Science*, 8, 44-56.
- McNamara, D.S. (Ed.: 2007). *Reading Comprehension Strategies: Theories, Interventions, and Technologies*. Erlbaum.
- Neale, M.D. (1989). *The Neale Analysis of Reading Ability* (2nd ed., revised), London: Macmillan Education.
- Oakhill, J., Yuill, N. & Donaldson, M. (1990). Understanding of "because" in skilled and less-skilled text comprehenders. *British Journal of Developmental Psychology*, 8, 401-410.
- Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117–75.
- Paris, S. G., & Jacobs, J. E. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development*, 55, 2083-2093.
- Paris, S.G. & Stahl, S. (2005). *Children's Reading Comprehension and Assessment*. Routledge.
- Pearce, D., Kerawalla, L., Luckin, R., Yuill, N. and Harris, A. (2005). The Task Sharing Framework: A generic approach to scaffolding collaboration and meta-collaboration in educational software. In C. Looi, D. Jonassem & M. Ikeda (Eds.), *Towards Sustainable and Scalable Educational Innovations Informed by the Learning Sciences*. IOS Press, (Singapore).
- Pearce, D. & Luckin, R. (2007). The principle of state expansion in task state-space navigation. In R. Luckin, K.R. Koedinger & J. Greer (Eds.), *Artificial Intelligence in Education: Building Technology-Rich Learning Contexts that Work*. Proceedings of AIED 13. IOS Press. Pp. 9-16.
- Piaget, J., & Inhelder, B. (1969). *The psychology of the child* (H. Weaver, Trans.). New York: Basic Books. (Original work published in 1966)
- Sheingold, K., Hawkins, J. and Char, C. (1990). 'I'm the thinkist, you're the typist': The interaction of technology and the social life of classrooms. In O. Boyd-Barrett and E. Scanlon, *Computers and Learning*, UK: Addison-Wesley.
- Siegler, R.S. (2005). Children's learning. *American Psychologist*, 60, 769-778.
- Sternberg, R. J., & Berg, C. A. (Eds.). (1992). *Intellectual development*. New York: Cambridge University Press.

- Teasley, S. and Roschelle, S. (1993). Constructing a joint problem space: the computer as a tool for sharing knowledge, in S. Lajoie and S. Derry, *Computers as cognitive tools*, Hillsdale: LEA.
- Vygotsky, L. (1978). *Mind in Society: The development of higher mental processes*. Cambridge, MA: Harvard University Press. (Works originally published 1930, 1933, 1935).
- Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *The Modern Language Journal*, 81, 470-481.
- Wells, G. & Chang-Wells, G.L. (1992) *Constructing knowledge together*. Portsmouth, NH: Heinemann.
- Wertsch, J. & Bivens, J.A. (1992). The social origins of individual mental functioning: Alternatives and perspectives. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 14, 35-44.
- Yuill (in press). The relation between ambiguity understanding and metalinguistic discussion of joking riddles in good and poor comprehenders: potential for intervention and possible processes of change. *First Language*.
- Yuill, N., & Joscelyne, T. (1988). Effect of organisational cues and strategies on good and poor comprehenders' story understanding. *Journal of Educational Psychology*, 80, 152- 158.
- Yuill, N. & Oakhill, J. (1992). *Children's Problems in Reading Comprehension: An Experimental Investigation*. Cambridge, UK: Cambridge University Press.

Figure 1 Format of sample SCoSS task

## Separate Control of Shared Space



- Simultaneous separate control
- Equal opportunity to participate
- Cannot over-ride partner's work
- Both need to click own 'agree' button
- Can see agreement and disagreement
- Can see own and other's working

Figure 2. Example of Reading Multiple Classification Task Items

chin	toe
chest	tooth
cheek	tongue
cheese	toffee
chocolate	toast
chip	tomato

Table 1. Study 1. Characteristics of three treatment groups  
(standard deviations in parentheses)

Condition	Girls	Boys	Age (mo)	Vocab./45	Pre-test accuracy*	Pre-test acc/speed x 10
SCoSS	14	10	103.2 (6.9)	36.1 (8.5)	5.49 (0.8)	.91 (.08)
Dual control	12	4	103.3 (8.4)	34.8 (9.7)	4.90 (1.0)	.68 (.07)
Individual	3	7	106.8 (6.7)	33.6 (9.7)	8.05 (1.4)	.51 (.06)

\* corrected for age to allow comparison with figures reported in text for post-test accuracy

Table 2. Study 2. Mean Age, Reading Ages and Pre-test Scores  
of Good and Poor Comprehenders (s.d.s in parentheses)

Group	Mean Age (yr:mo)	Neale Accuracy Age	Neale Comprehension Age	Pre-test Story recall (/30)	Pre-test Question Answering (/15)
poor	9;2 (3.5)	10;1 (14.4)	7;3 (8.0)	13.75 (4.5)	9.13 (2.8)
good	9;3 (4.3)	9;8 (18.9)	9;10 (18.6)	15.88 (5.2)	12.63 (1.6)
Difference	F<1	F<1	F (1, 14) = 18.8, p<.001, eta sq .57	F<1	F (1, 14) = 9.2, p<.01, eta sq .40