The Roles of Private Speech and Inner Speech in Planning in Middle Childhood: Evidence from a Dual Task Paradigm

Jane S. M. Lidstone, Elizabeth Meins, and Charles Fernyhough

Durham University, UK

Author Note

All authors are at the Psychology Department, Durham University.

Acknowledgements: The study was conducted in partial fulfilment of Jane Lidstone's PhD, and was funded by a studentship from the Economic and Social Research Council, with additional funds from Durham University. Many thanks to Chrisoph Kaller for help with Tower of London problem selection, Joyce Ching for coding for inter-rater reliability, and Bob Metcalfe for creating the foot-tapping soundtrack. Special thanks to the participants for their enthusiastic participation and the school staff who made the study possible.

Address for correspondence: Jane Lidstone, Psychology Department, Durham University, Science Laboratories, South Road, Durham, DH1 3LE, UK; Tel: +44(0)1928 73 01 07; Fax: +44(0)191 334 3241; Email: j.s.m.lidstone@durham.ac.uk

The manuscript is 8487 words

The Roles of Private Speech and Inner Speech in Planning in Middle Childhood: Evidence from a Dual Task Paradigm

Abstract

Children often talk themselves through their activities, producing private speech, which is internalized to form inner speech. The present study assessed the effect of articulatory suppression (which suppresses private and inner speech) on Tower of London performance in 7- to 10-year-olds. Experiment 1 ($N = 30$) showed no effect of articulatory suppression on performance with the standard Tower of London procedure; we interpret this in terms of a lack of planning in our sample. Experiment 2 ($N = 30$) used a modified procedure in which participants were forced to plan ahead. Performance in the articulatory suppression condition was lower than in the nonverbal control condition, consistent with a role for self-directed (private and inner) speech in planning. On problems of intermediate difficulty, participants producing more private speech in the nonverbal control condition showed greater susceptibility to interference from articulatory suppression than their peers, suggesting that articulatory suppression interfered with performance by blocking self-directed (private and inner) speech.

Keywords: Tower of London; Tower of Hanoi; children; dual task paradigm.

The Roles of Private Speech and Inner Speech in Planning in Middle Childhood: Evidence from a Dual Task Paradigm

Vygotsky (1934/1987) saw higher mental functions such as flexible goal-directed thought as being founded upon the experience of participating in dialogue around joint activity. The ability to regulate one's own thought and behavior is seen as emerging from the experience of taking part in interactions in which adult and child use speech to direct each others' thought and behavior. When children first use speech to direct their own thought and behavior, they are said to be producing *private speech*. Private speech describes utterances spoken aloud that appear to serve a self-regulatory function rather than a communicative function: They are self-directed, and often take the form of self-guiding comments. Private speech is mainly found in preschoolers, but can appear in middle childhood and adulthood, when it is likely to take the form of more covert muttering and whispering (see Winsler, 2009). It is thought that this shift towards covertness reflects the gradual internalization of private speech to form inner speech, or silent verbal thought (Vygotsky, 1934/1987). Private speech and inner speech together are hereafter referred to as self-directed speech (Figure 1).

Self-directed speech has been implicated in the performance of problem-solving tasks, some spatial working memory tasks, and executive functions, in studies which will be described below. Some of this evidence comes from studies relating private speech production to task performance. A cognitive task is thought to be reliant on self-directed speech if private speech production predicts either concurrent or future performance in children. For example, Winsler, Diaz, and Montero (1997) had preschoolers perform a selective attention task, each trial of which required them to determine which of two perceptual dimensions (shape or color) was shared by two pictures, and then to select, from a group of alternatives, the answer card that represented the shared dimension. After receiving guidance from an experimenter, children were more likely to succeed if they used private speech than if they were silent. Similarly, Behrend, Rosengren, and Perlmutter (1989, 1992) found that preschoolers' private speech production during spatial problem-solving tasks correlated with both their concurrent and future performance of those tasks.

However, there are a number of problems with looking at private speech-performance relations to speak to whether or not tasks are reliant on self-directed speech. One is that private speech production shows a positive or curvilinear relation with task difficulty, and if this is not taken into account private speech-performance relations can be missed (see Fernyhough & Fradley, 2005; Frauenglass & Diaz, 1985). Even when they are found, the difficulty with a non-experimental design is that it leaves open the question of whether private speech is useful for or merely happens to accompany successful cognitive performance.

An approach that avoids these problems is to use the dual task paradigm to assess the effect of preventing self-directed speech. The experimental design allows researchers to investigate whether or not self-directed speech has a causal role in cognitive performance. Researchers can prevent the use of self-directed speech by asking participants to engage in an *articulatory suppression* task concurrently with the primary task on which performance is being assessed. Articulatory suppression can take the form of repeating a word, repeating a well-learned sequence of words like the months of the year, or shadowing prose heard while completing the primary task. (Articulatory suppression is usually referred to as suppressing "inner speech," but of course it interferes with private speech as well.) If performance of the primary task relies on self-directed speech, it should be significantly impaired by articulatory suppression. The performance of several cognitive tasks is vulnerable to articulatory suppression in children and adults, including tasks tapping spatial working memory (Ang & Lee, 2008), and task-switching (Whitehouse, Maybery, & Durkin, 2006) in children, and tasks tapping spatial reasoning (Kim, 2002), cognitive flexibility (Baldo et al., 2005) and task-switching performance (Baddeley, Chincotta, & Adlam, 2001) in adults.

In many of these studies, performance in the articulatory suppression condition was compared to performance in a control condition with no secondary task. However, as Emerson and Miyake (2003) point out, the effect of articulatory suppression in some cases might be wholly attributed to the general demands of performing two tasks simultaneously. To guard against this possibility, a nonverbal secondary task, such as foot-tapping, can be included in the control condition. If the articulatory suppression task is to say *a b c* once every metronome beat, the control task would be to tap one’s foot once every metronome beat. If the articulatory suppression task is verbal shadowing, an appropriate control condition could involve shadowing a rhythm by foot-tapping. Foot-tapping is thought to be a good control task because, like articulatory suppression, it incorporates a motor component, and it involves an attentional component that is similar to that of articulatory suppression (Robbins et al., 1996). Its suitability was tested by Emerson and Miyake (2003), who found that, on a visual task assumed to be completely nonverbal (the Identical Pictures Test), articulatory suppression and foot-tapping affected adults’ performance equally. Foot-tapping is now included in the control conditions of studies assessing the effect of articulatory suppression on task-switching. They show that articulatory suppression impairs performance to a greater extent than does foot-tapping (e.g., Emerson & Miyake, 2003; Liefooghe, Vandierendonck, Muyllaert, Verbruggen, & Vanneste, 2005; Saeki, Saito, & Kawaguchi, 2006), suggesting that task-switching relies on self-directed speech. Other research has revealed effects of articulatory suppression on spatial reorientation (that is, the ability to integrate geometric and landmark cues in order to reorient oneself in space after disorientation; Hermer-Vasquez, Spelke, & Katsnelson, 1999) and face learning (Nakabayashi & Burton, 2008) compared to tapping control conditions.

One function that has received surprisingly little attention in research on self-directed speech is that of planning. Planning is surely one of the most common human mental activities, and is at the very core of goal-directed behavior (Cohen, 1996). According to Vygotskian theory, self-directed speech has a special role in planning. Vygotsky's own studies suggested that one of the most significant developments of private speech in the preschool years is that it takes on a planning function. Upon discovering the planning function of speech, he argues, "[children's] psychological field changes radically. A view of the future is now an integral part of their approaches to their surroundings" (Vygotsky, 1930-1935/1978, p. 29). He argues that speech (or "verbal signs") is helpful in acting as a barrier between impulsive and actual behavior. Thus, "the inclusion of signs ... creates the conditions for the development of a single system that includes effective elements of the past, present, and future. This emerging psychological system in the child now encompasses two new functions: intentions and symbolic representations of purposeful action" (pp. 36-37).

On this view, self-directed speech should be particularly useful for the performance of tasks requiring planning in childhood. The gold standard planning tasks are the Tower of Hanoi, and its more commonly used adaptation, the Tower of London (Shallice, 1982). The Tower of London consists of three different-colored disks, arranged on three pegs that can hold one, two, and three disks respectively (Figure 2). Participants attempt to transform one configuration into another by moving one disk at a time. Planning is required because participants must complete the task in the smallest number of moves possible.

To our knowledge, there are in the extant literature four studies with results that speak directly to whether or not self-directed speech is involved in Tower of London or Tower of Hanoi performance; they all relate to the Tower of London. The first to be considered here is a study of the private speech of 5- and 6-year-olds while completing the Tower of London (Fernyhough & Fradley, 2005). The authors found that children producing more private speech completed the task more quickly and accurately than children who produced less. This was partially replicated by Al-Namlah, Fernyhough, and Meins (2006), who found a negative association between private speech production and the time taken to complete Tower of London problems in their sample of 4- to 8-year-olds. These findings are consistent with the idea that successful planning requires self-directed speech in early childhood.

Wallace, Silvers, Martin, and Kenworthy (2009) reported the effect of articulatory suppression on Tower of London performance in a group of typically developing adolescents (12- to 19-year-olds) – the control group in a study of "inner speech" in autism. The participants completed Tower of London problems, alternately with and without articulatory suppression. Under articulatory suppression, the typically developing participants took significantly more moves to complete the problems than they did without articulatory suppression. The authors interpreted the results to mean that inner speech supported performance in their sample of typically developing adolescents. Because there was no control secondary task, however, the results are open to the alternative interpretation mentioned above: that the effect of articulatory suppression could be wholly attributed to general dual task effects.

The fourth study (Phillips, Wynn, Gilhooly, Della Sala, & Logie, 1999) also did not include a secondary task in the control condition, but this was less problematic for our purposes given the pattern of results. The participants, young adults aged 18 to 25 years, completed Ward and Allport's (1997) five-disk Tower of London with and without articulatory suppression. Articulatory suppression was not detrimental to performance accuracy (defined in terms of the number of excess moves) for problems of any level of difficulty. The effect of articulatory suppression was only to speed up performance, although its effect in reducing planning times (time to first move) mainly occurred for the most difficult problems, which were too complex to be planned in full even with no secondary task (Phillips, Wynn, McPherson, & Gilhooly, 2001).

Thus there are three studies (Al-Namlah et al., 2006; Fernyhough and Fradley, 2005; Wallace et al., 2009) suggesting a role for self-directed speech in planning, but none used dual task methodology with a dual task control condition. The other study (Phillips et al., 1999) is perhaps the most conclusive of these four investigations in terms of our question of whether self-directed speech is important for planning, suggesting it is not. The lack of articulatory suppression interference on planning accuracy in adults does not preclude the possibility of finding an effect earlier in development, however. We predicted that planning would be largely dependent on self-directed speech in middle childhood. To test this hypothesis was the principal aim of the present study. Specifically, we predicted that Tower of London performance would be impaired under articulatory suppression relative to a control condition with a foot-tapping task.

The second hypothesis was that the detrimental effect of articulatory suppression on performance would be larger for children whose performance relied on self-directed speech to a greater extent, as evidenced by more frequent private speech production in the tapping condition. In this way we hoped to provide further evidence that articulatory suppression has its detrimental effect on primary task performance by interfering with self-directed speech, rather than through general dual task demands. We expected to find a relation between private speech production and interference by articulatory suppression only for problems for which private speech was useful. Like Fernyhough and Fradley (2005), who looked at private speech-performance relations, we expected the utility of private speech to be moderated by task difficulty. For the easiest problems, we expected speech to be mainly fully internalized, meaning that private speech production would be a good indicator of the extent to which performance was reliant on self-directed speech for intermediate and difficult problems only.

For the most difficult problems, beyond the children's ability range, private speech was predicted to be ineffective for improving performance in the control condition. We therefore predicted a positive relation between private speech production and interference by articulatory suppression only for problems of intermediate difficulty. Inner speech, on the other hand, was predicted to be useful for the easiest problems. As articulatory suppression interferes with both private speech and inner speech, we expected articulatory suppression to be detrimental to performance on the easiest problems as well as those of intermediate difficulty.

Experiment 1

In Experiment 1, we sought to test the two hypotheses described above – that planning would be disrupted by articulatory suppression in middle childhood, and that the amount of private speech produced in the foot-tapping condition would correlate positively with articulatory suppression interference for problems of intermediate difficulty.

Method

Participants. The participants were 30 typically developing children (13 boys), recruited from and tested in mainstream state schools in the North-East of England. The mean age of the children was 9 years; 1 month (*SD* 0;9, range 7;11 – 10;5). No participant had a learning or neurological disorder according to teacher report. All had active written parental consent to participate, and were free to withdraw at any time.

Materials. The Tower of London consisted of two wooden frames, each with three colored disks (Figure 2). A camcorder recorded the testing sessions. A program on a laptop computer, connected to a foot pedal, was used for the tapping task: It produced sounds to allow participants to monitor their foot-tapping performance (see below). The pedal was mounted on a wooden platform, which incorporated an adjustable foot rest.

There were two sets of 13 Tower of London problems – 10 experimental problems plus 3 practice problems – one set for each condition. The problem sets were identical except that the colors of the disks were swapped around; that is, the sets were isoforms of each other. The practice trials were 1-, 2-, and 3-move problems, none of which were duplicated in the experimental problem set. The experimental problem set consisted of two 2-move problems, three 3-move problems, three 4-move problems and two 5-move problems. No problem appeared in the same problem set twice. Although the minimum number of moves is not the only aspect of Tower of London problems that influences task difficulty (Kaller, Unterrainer, Rahm, & Halsband, 2004), it is hereafter used as a rough guide to the difficulty level of the problems.

Procedure. The participants completed the two dual task conditions in a single session. The order of conditions was counterbalanced so that the two groups – those receiving the tapping condition first and those receiving the articulatory suppression condition first – did not differ in gender composition or chronological age.

The participants were told that their job was to make the two puzzles look the same, by moving one disk at a time, and that they would “need to plan ahead” to do so in the minimum number of moves. The problems were presented in order of increasing difficulty, and the participants were told the number of moves they should use to solve each problem. Participants received a sticker for each problem they solved in the minimum number of moves and another for each problem that was completed with no secondary task errors.

The secondary tasks, repeating the word *Monday* (articulatory suppression) and foot-tapping (control), were demonstrated by the experimenter, who performed them at a rate of one response per second. Participants then practised the secondary tasks with the Tower of London practice trials.

In the tapping condition, each tap was accompanied by a beep. The beeping was intended to serve as an aural reminder of the task. If there was an error, defined as a gap between taps of 2.0 seconds – equal to missing one tap – there was a warning sound, which ceased when tapping was recommenced.

The aural reminder of the articulatory suppression task was the experimenter's articulation of *Monday* in time with that of the participants. If a participant made an error, defined as a missed *Monday*, the experimenter reminded her to recommence by uttering her name.

Scoring and analysis. Two commonly-used measures of Tower of London performance are the number of excess moves (i.e., the difference between the number of moves taken to solve a problem and the minimum number of moves), and whether or not a problem was solved in the minimum number of moves (Berg & Byrd, 2002). The latter was more compatible with the instructions given to participants (which were designed to focus their attention on the need for careful planning) and this measure had the advantage of rendering the results of Experiments 1 and 2 comparable. The primary outcome measure for each trial was therefore whether or not it had been solved in the minimum number of moves. A trial was considered to have ended after the first incorrect move, as incorrect sequences of moves often ended in an impasse and participants were stopped by the experimenter. The secondary performance measure, time taken to complete the problems, was therefore measured only for correctly-solved problems. The third measure of performance on each trial was whether or not one or more secondary task errors had been made before the end of the trial.

The participants' speech in the tapping condition was coded from the video recordings. Private speech was defined as any speech that did not meet the criteria to be regarded as social speech (Winsler, Fernyhough, McClaren, & Way, 2005). Social speech was defined as any full volume speech intended for communication with the experimenter. Communicative intent was identified where the participant involved the experimenter (through physical contact, gaze direction, etc.), during or within two seconds of an utterance (Winsler et al., 2005). The frequency of social speech was negligible so it is not reported.

Private speech is traditionally coded according to Berk (1986) as *Level 1* (task-irrelevant private speech), *Level 2* (task-relevant externalized private speech), or *Level 3* (presumably task-relevant external manifestations of inner speech, including inaudible muttering and whispering, and silent, verbal lip movements). However, the frequency of task-irrelevant private speech was negligible, and the internalization level of private speech was not relevant to our hypothesis. Therefore each trial was coded as containing or not containing task-relevant private speech (Levels 2 and 3 together). A trial-based metric was chosen as rate-based metrics (such as utterances per minute) risk confounding general verbosity with the degree of dependence on private speech (Winsler et al., 1997), especially where it is not possible to control for verbosity by partialing out social speech production (see Winsler et al., 2005). A second researcher independently coded 20% (six) of the recordings. Inter-rater agreement for the presence of private speech was $\kappa = .87$.

The frequency of private speech was a function of the percentage of trials containing private speech at each difficulty level. In order that the results would be directly comparable to those of Experiment 2, which had an equal number of problems of each level of difficulty, we applied a weighting to the problems of Experiment 1. Private speech rates were weighted so that each difficulty level was represented equally. For example, without weighting, a child producing private speech during the performance of neither 2-move problem, all three 3-move problems, all three 4-move problems and neither 5-move problem would score 60%. With weighting, the rate would be the mean of 0%, 100%, 100% and 0%, which is 50%.

The principal measure of task performance in each condition was the percentage of problems solved in the minimum number of moves, with the weighting system applied. Other measures of task performance were the time taken to complete correctly-performed problems, and the percentage of problems containing a secondary task error. The same weighting system applied to all performance variables. For response times, the mean time was found for correctly-performed 2-, 3-, 4-, and 5-move problems separately, and then the grand mean was taken. The weighting system did not change the results of any statistical test: It simply made the accuracy of Tower of London performance more comparable across experiments.

Parametric statistics were used throughout. Although one of the variables (the proportion of problems containing a secondary task error) was positively skewed, parametric statistics were robust because the distribution was the same in each condition, the variances were similar, and there were more than 20 degrees of freedom (Tabachnick & Fidell, 2007). Difficulty level was included in analyses of performance accuracy, as the effect of articulatory suppression was expected to vary with this. Time taken to complete correctly-performed trials and the percentage of trials with a secondary task error were analyzed with 2 $\times$ 2 (Condition [articulatory suppression, tapping] $\times$ Condition Order [articulatory suppression first, tapping first]) repeated measures ANOVAs.

Results and Discussion

In the control condition, the mean (*SD*) percentage of trials with private speech for 2-, 3-, 4-, and 5-move problems was 5 (15), 9 (15), 6 (13), and 10 (24), respectively. The mean proportion of all trials with private speech in the control condition was 7% (*SD* 11, range 0 to 50).

Performance in terms of the percentage of problems solved correctly is shown in Figure 3. A 2 $\times$ 2 $\times$ 4 (Condition [articulatory suppression, tapping] $\times$ Condition Order [articulatory suppression first, tapping first] $\times$ Difficulty Level [2-, 3-, 4-, 5-moves]) repeated measures ANOVA was used to predict the percentage of problems solved correctly. There was a main effect of difficulty level, $F(3, 84) = 75.60$, $p < .001$. Within-subjects contrasts revealed a linear trend, $F(1, 29) = 184.96$, $p < .001$, with performance decreasing with increasing difficulty level. There was neither a main effect of condition, $F < 1$, nor a Condition × Difficulty Level interaction, $F < 1$. No other effects approached significance (all *p*s > .25).

In terms of the time taken to complete correctly-performed trials, there were no effects of condition, $F < 1$, or condition order, $F < 1$, and there was no Condition × Condition Order interaction, $F(1, 28) = 1.50$, $p = .23$. Thus, there was no difference in the time taken to complete correctly-performed trials in the articulatory suppression condition, $M = 8.7$ s, $SD = 2.3$, and compared with the control condition, $M = 8.3$ s, $SD = 3.4$.

In terms of the percentage of trials with a secondary task error, there was no main effect of condition order, $F(1, 28) = 1.06$, $p = .31$. There was a marginally significant main effect of condition, $F(1, 28) = 4.32$, $p = .05$, with more errors in the articulatory suppression condition, $M = 11.9\%$, $SD = 16.1$, than in the control condition, $M = 5.6\%$, $SD = 9.7$. However, this was modified by a Condition × Condition Order interaction, $F(1, 28) = 6.87$, $p = .01$. Follow-up *t*-tests showed that, amongst those receiving articulatory suppression first, the secondary task error rate was higher in the articulatory suppression condition, $M = 17.8\%$, $SD = 19.6$, than in the tapping condition, $M = 3.3\%$, $SD = 6.9$, $t(14) = 2.79$, $p = .01$. Among participants receiving the tapping condition first, there was no difference in secondary task error rate between the articulatory suppression condition, $M = 6.1\%$, $SD = 9.0$, and the tapping condition, $M = 7.8\%$, $SD = 11.8$, $t(14) = 0.50$, $p = .62$.

In sum, there were more secondary task errors in the articulatory suppression condition than in the tapping condition, but this was limited to participants receiving the articulatory suppression condition first. Articulatory suppression had no effect on Tower of

London performance, suggesting the latter was not dependent on self-directed speech. Therefore the second hypothesis was not tested.

In response to these results, the video recordings were re-examined. We recorded the time taken to initiate the first move for each trial – the planning time (see Berg & Byrd, 2002) – in the control condition. We chose the control condition rather than the articulatory suppression condition because the former is the condition in which planning is theoretically uninhibited. One participant's video recording was lost after a technical problem with the camcorder, so the results on planning times relate to 29 participants.

This analysis revealed planning times to be very short, $M = 3.1$ s, $SD = 1.1$. In addition, planning times did not increase with the difficulty level of the problems: Means in seconds (with standard deviations in parentheses) for 2- through 5-move problems were 2.7 (1.5), 3.5 (1.6), 2.9 (1.5), and 3.4 (2.5) respectively. The lack of relation between planning times and difficulty level was confirmed by a repeated measures ANOVA: There was no effect of difficulty level on planning times, $F(3, 84) = 1.90$, $p = .14$, and the within-subjects contrasts indicated no significant linear trend, $F(1, 28) = 0.75$, $p = .39$.

These planning times are markedly shorter than those in Phillips et al. (1999), which averaged around 15 seconds in the control condition (planning times are not reported in Fernyhough & Fradley, 2005, Al-Namlah et al., 2006, or Wallace et al., 2009). In addition, unlike in Phillips et al., planning time did not increase with trial difficulty. From this analysis, we concluded that performance in the present study was not dependent on self-directed speech perhaps because the procedure was not effective in eliciting planning. We therefore conducted a second experiment using another Tower of London procedure in which participants were forced to plan ahead. Instead of asking participants to move the disks to make the configurations match, we asked them to mentally plan the moves, to tell the experimenter the minimum number of moves it would take to make the configurations match, and then to demonstrate the moves they had planned. The original *How many moves* procedure, in which participants did not have to additionally demonstrate the moves, was created by Owen et al. (1995) for use with adults, and has been used in several subsequent neuroimaging studies (e.g., Baker et al., 1996; Boghi et al., 2006). The predictions were as for Experiment 1.

Experiment 2

Method

Participants. The participants were 30 typically developing children (16 boys), recruited in the same way as the participants in Experiment 1. No child participated in both experiments. The mean age was 9 years; 4 months (*SD* 0;9, range 7;10 – 10;8).

Materials. All materials were as above, except that in Experiment 2 there were 8 instead of 10 experimental problems per condition. The number of problems was changed so that there could be an equal number of problems of each difficulty level, i.e., two 2-move, two 3-move, two 4-move, and two 5-move problems. This was to ensure that guessing *3 moves* or *4 moves* would not be reliably more effective than guessing *2 moves* or *5 moves*.

Procedure. In each condition, the problems were administered in a different pseudo-randomized order. Pilot work indicated that the children would need more than the three practice problems provided in Experiment 1, so they completed one practice set of eight problems, before completing the dual task conditions in a second session about a week later. The order of dual task conditions was counterbalanced as before.

As the Tower of London was introduced, the participants were asked "to imagine moving the disks around, one at a time, and tell [the experimenter] how many moves it would take to make [the start state] look like [the end state]." For the experimental problems, participants were just asked "How many moves?" for each trial. Unlike in previous studies using this version of the Tower of London (Baker et al., 1996; Boghi et al., 2006; Owen et al., 1995), the participants were asked to demonstrate the moves after telling the experimenter the number of moves they had planned.

The details of the secondary tasks were as above. They were performed only during the planning phase, the period between the start of the trial and the verbal response. Similarly, only the planning phase was coded for private speech.

Scoring and analysis. A Tower of London problem was scored as correct if the participant both named and correctly demonstrated the minimum number of moves required to make the start and end states match. The response time – the time from presentation of a problem to the verbal numerical response – was also recorded. As in Experiment 1, trials were coded dichotomously on the basis of whether or not a secondary task error had been made.

Private speech was coded as in Experiment 1. Inter-rater reliability was $\kappa = .90$. Video recordings were unavailable for one participant because of a technical problem with the camcorder, so the response time and private speech data relate to 29 participants.

No weighting system was used in Experiment 2 as there was an equal number of problems at each level of difficulty. As in Experiment 1, the principal measure of task performance was the percentage of problems solved in the minimum number of moves. Other measures of performance were response times, and the percentage of problems containing a secondary task error. Analyses were performed as in Experiment 1.

Results and Discussion

As for Experiment 1, a 2 × 2 × 4 (Condition [articulatory suppression, tapping] × Condition Order [articulatory suppression first, tapping first] × Difficulty Level [2-, 3-, 4-, 5-moves]) repeated measures ANOVA was used to predict the percentage of problems solved correctly. There was a main effect of condition, $F(1, 29) = 10.55$, $p = .003$; performance was impaired in the articulatory suppression condition compared to the control condition (Figure

3). There was a main effect of difficulty level, $F(3, 87) = 86.80$, $p < .001$, but no Condition × Difficulty Level interaction, $F < 1$, with articulatory suppression affecting performance at all difficulty levels equally. No other effects approached significance (*p*s > .15).

Response times and secondary task error rates were analyzed using 2 × 2 (Condition [articulatory suppression, tapping] × Condition Order [articulatory suppression first, tapping first]) repeated measures ANOVAs. In terms of response times, there was no effect of condition, $F(1, 27) = 1.31$, $p = .26$. Thus, the response times in the articulatory suppression condition, $M = 13.4$ s, $SD = 4.5$, did not differ from those in the tapping condition, $M = 14.3$ s, $SD = 3.8$. There was no main effect of condition order, $F < 1$, but the Condition × Condition Order interaction approached significance, $F\ (1, 27) = 3.52$, $p = .07$. Follow-up *t*-tests showed that, among participants receiving the articulatory suppression condition first, there was no difference in response times between the articulatory suppression condition, $M = 14.1$ s, $SD = 5.4$, and the tapping condition, $M = 13.4$ s, $SD = 3.1$, whereas, among those receiving the tapping condition first, response times were shorter in the articulatory suppression condition, $M = 12.6$ s, $SD = 3.5$, than in the tapping condition, $M = 15.3$ s, $SD = 4.3$, $t(13) = 2.63$, $p = .02$. The participants who received the tapping condition first thus exhibited an improvement in their response times, unlike those receiving the articulatory suppression condition first.

In terms of the percentage of trials with a secondary task error, there was a main effect of condition, $F(1, 28) = 7.76$, $p = .01$, with more articulatory suppression errors, $M = 19.6\%$, $SD = 17.3$, than tapping errors, $M = 8.8\%$, $SD = 12.8$, as in Experiment 1. There was a marginally significant main effect of condition order, $F(1, 28) = 3.92$, $p = .06$, which was not modified by a Condition × Condition Order interaction, $F < 1$. The proportion of trials with secondary task errors was lower among those receiving articulatory suppression first than among those receiving tapping first in both the articulatory suppression condition, $M =$

14.1%, *SD* = 16.3, vs. *M* = 25.0%, *SD* = 17.0, and the tapping condition, *M* = 6.6%, *SD* = 12.4 vs. *M* = 10.9%, *SD* = 13.3. Perhaps receiving the articulatory suppression condition first biased the participants toward allocating more attentional resources to the secondary task. Considering the deleterious effect of articulatory suppression on primary task performance, and the fact that participants were rewarded equally for perfect articulatory suppression performance and perfect Tower of London performance (with one sticker for each), this would be the optimum strategy in the articulatory suppression condition, and it was presumably carried over to the tapping condition by participants receiving this second. Overall, though, secondary task performance was poorer in the articulatory suppression condition than the tapping condition.

In sum, Tower of London performance appeared to be dependent on self-directed speech in Experiment 2. Tower of London performance and secondary task performance were lower in the articulatory suppression condition than in the control condition. The fact that the effect of articulatory suppression on primary task performance did not vary by difficulty level is discussed below.

To find out whether the effect of condition on Tower of London performance in Experiment 2 was significantly different from that in Experiment 1, we combined the results into a single 2 × 2 × 4 (Experiment [Experiment 1, Experiment 2] × Condition [articulatory suppression, tapping] × Difficulty Level [2-, 3-, 4-, 5-moves]) ANOVA predicting the percentage of problems solved correctly. There was a main effect of condition, $F(1, 58) = 6.49$, $p = .01$, which was modified by a Condition × Experiment interaction, $F(1, 58) = 5.64$, $p = .02$. Results shown above indicate this was due to an effect of dual task condition in Experiment 2 but not in Experiment 1. There was a main effect of difficulty level, $F(3, 174) = 159.76$, $p < .001$; no other effects approached significance (defined as $p < .10$).

The absence of a main effect of Experiment in the above ANOVA was perhaps surprising, given that little planning took place in Experiment 1 – a factor we would expect to result in lower success rates in Experiment 1 than Experiment 2. Comparison of the control conditions confirmed that success rates on the two Tower of London versions did not differ: The mean proportion of problems solved correctly in the control condition of Experiment 1 was 66.6% (*SD* = 13.9), and in Experiment 2 it was 67.1% (*SD* = 18.4), $t(58) = 0.36$, $p = .72$. Possible explanations of the equal success rates in the two experiments are considered in the General Discussion.

Next, our attention turned to the private speech results of Experiment 2's control condition: Would there be a correlation between articulatory suppression interference and private speech production for problems of intermediate difficulty? Private speech was produced during 47% of the trials on average (*SD* 39, range 0 to 100). The mean (*SD*) percentages of trials with private speech for 2-, 3-, 4-, and 5-move problems were 34 (42), 41 (44), 50 (46), and 60 (45), respectively. The percentage of trials with private speech appeared to increase with difficulty level, but note that trial duration also increased with difficulty level (data not shown). Our measure of articulatory suppression interference was the percentage of trials correct in the articulatory suppression condition minus the percentage of trials correct in the control condition. Thus, a positive figure indicated poorer performance in the articulatory suppression condition than in the control condition. We used this *difference score* in line with previous research on individual differences in articulatory suppression interference (Lidstone, Fernyhough, Meins, & Whitehouse, 2009), but we also calculated a measure of articulatory suppression interference based on *residual scores*, by partialing the control condition performance from the articulatory suppression condition performance. As the correlations using these residual scores produced exactly the same pattern as those using the difference scores, only the latter are shown below.

The correlations between articulatory suppression interference and the proportion of trials with private speech in the control condition for 2-, 3-, 4- and 5-move problems (with 29 degrees of freedom) were .07, -.10, .47, and .21, respectively (Pearson's correlation coefficients). The largest of these was statistically significant ($p = .01$). As expected, then, the relation between private speech production and interference by articulatory suppression was found for problems of intermediate difficulty only. These results indicate that children who produced more private speech on 4-move problems experienced greater interference from articulatory suppression on these problems, and therefore that articulatory suppression has its detrimental effect on Tower of London performance by suppressing self-directed speech.

General Discussion

The principal aim of the study was to investigate, using the dual task paradigm, whether or not planning relies on self-directed speech in middle childhood. Experiment 1 showed no effect of articulatory suppression on performance of the standard Tower of London. However, this was interpreted as being due to a lack of planning in our sample. Experiment 2 showed that, when participants were forced to plan ahead, suppressing self-directed speech was detrimental to Tower of London performance. The results of Experiment 2 support those of Fernyhough and Fradley (2005), Al-Namlah et al. (2006), and Wallace et al. (2009) in suggesting that planning is achieved with the aid of self-directed speech. These findings are consistent with Vygotsky's (1930-1935/1978; 1934/1987) ideas on the role of speech in cognition, and suggest planning can be considered to be largely verbally mediated in middle childhood. The results are consistent with the view that cognition undergoes a domain-general shift towards verbal mediation during early childhood (Al-Namlah et al., 2006; Fernyhough, 2008).

The reason for the lack of planning in Experiment 1 is unknown. There were instructions to plan ahead, and the participants were told how many moves each problem should take and were only rewarded (with stickers) if they solved the problems in the specified number of moves, emphasizing that reaching the goal state in more moves than necessary did not constitute a correct answer. In retrospect, our intuition is that starting each session with 2- and 3-move problems might have contributed to the lack of planning for two reasons. First, the participants perhaps did not get into the habit of planning as the first five problems (of 2 and 3 moves) could quite easily be solved correctly with little advance planning. Second, by the time the children reached the more difficult problems, they perhaps felt comfortable making mistakes, having achieved a good success rate in the earlier part of the session.

In light of the fact that little planning took place in Experiment 1, it is perhaps surprising that control condition performance equalled that of Experiment 2's control condition.[1] We propose that this can be explained in terms of the fact that, in Experiment 1, the problem-solving activity of the participants was in effect carefully "scaffolded," in that the problems were presented in exact order of increasing difficulty. In terms of performance levels, the helpful effect of this scaffolding probably counteracted the detrimental effect of reducing planning.

We have characterized the procedure used in Experiment 2 as requiring more planning than that used in Experiment 1, but the procedure of Experiment 2 undoubtedly drew more heavily on working memory as well. In our view, however, any concept of planning that requires no memory is of limited value. Memory is surely vital to planning, because tentative and incomplete plans need to be held in mind while they are evaluated and revised (Cohen, 1996). On this view, to conceptualize the version of the Tower of London used in Experiment 2 as requiring a greater degree of planning is appropriate. In any case, the finding that children's performance of this seemingly spatial task was dependent on self-directed speech still stands.

The fact that the Tower of London procedure which elicited little planning was equally affected by articulatory suppression and foot-tapping is reminiscent of a finding from a previous study relating to the general dual task demands of articulatory suppression and foot-tapping in adults (Emerson & Miyake, 2003). As mentioned in the Introduction, these authors reported that, on a visual task assumed to be nonverbal, articulatory suppression and foot-tapping affected adults' performance equally. The present study's Experiment 1 results, indicating equal effects of the two secondary tasks, could be interpreted as preliminary evidence that the secondary tasks exert equivalent dual task demands also in children, and that any deleterious effect of articulatory suppression can be attributed to its effect of suppressing self-directed speech. However, the meaning of the trend towards more secondary task errors in the articulatory suppression condition is unclear, and perhaps counters that claim.

Clearer evidence on the issue of whether articulatory suppression has its effect specifically by blocking self-directed speech comes from the combination of the dual task paradigm with the private speech results. As expected, children who produced more private speech during 4-move problems evidenced a greater difference in performance between the articulatory suppression and control conditions. This suggests that the difference in performance between the dual task conditions related to the fact that, in the articulatory suppression condition, self-directed speech was suppressed.

Although we predicted that the relation between private speech and articulatory suppression interference would exist only for problems of intermediate difficulty, we predicted that there would be an effect of articulatory suppression for easy and intermediate problems, but not for the most difficult problems. The rationale was that the most difficult problems would be beyond the children's ability range and therefore private speech (and inner speech) would not be as useful as for the easier problems. In fact, the effect of articulatory suppression did not vary by difficulty level in Experiment 2. Perhaps 5-move problems were within the range at which private speech would improve performance; the correlation between private speech production and articulatory suppression interference for 5-move problems was positive (.21), though not statistically significant.

Limitations of the present study include the fact that the reason for the paucity of planning in Experiment 1 is unknown. Although a certain amount of planning probably occurred "online" (while moving the disks), the relation between this and advance planning is unknown (see Berg & Byrd, 2002). Similarly, the precise nature of planning as measured by the *How many moves* version of the Tower of London is somewhat underspecified. Owen et al.'s (1995) *How many moves* version, like the original Shallice's (1982) original procedure, is sensitive to frontal lobe lesions (Owen et al.), and comparison of functional neuroimaging studies shows that it activates the same neural network as Shallice's original procedure (Boghi et al., 2006), including the motor and prefrontal areas associated with planning (Baker et al., 1996). However, to our knowledge no study has directly compared the versions, and so it is not possible to go into detail about how they might differ, save for the observation that the *How many moves* version is likely to involve a larger memory component.[2] Such studies might prove valuable in the future.

Future research could also look at whether or not the reliance of planning on self-directed speech decreases between childhood and adulthood, as suggested by the present study in combination with that of Fernyhough and Fradley (2005), Al-Namlah et al. (2006), and Wallace et al. (all of which found effects in children or adolescents) and Phillips et al. (1999, which found no detrimental effect of articulatory suppression in adults). The difference between Phillips et al.'s and the others' findings might be explained in terms of the different ages of the participants. Alternatively, it might be an artifact of the differing task demands of the five-disk and three-disk versions of the Tower of London (Berg & Byrd, 2002). To see whether adults show an effect of articulatory suppression on the three-disk version might therefore be informative.

The present study is, to our knowledge, the first study investigating planning and indeed any executive function in children by documenting the effect of articulatory suppression relative to a control dual task condition. The results were clear: that planning is dependent on self-directed speech in middle childhood. We suggest that the dual task paradigm is a useful tool for the investigation of self-directed speech in childhood, particularly when used in conjunction with the observation of private speech.

Footnotes

[1] We thank an anonymous reviewer for drawing this to our attention.

[2] We do not consider the present study to have compared the versions because the Experiment 1 procedure did not elicit planning as it has in previous studies.

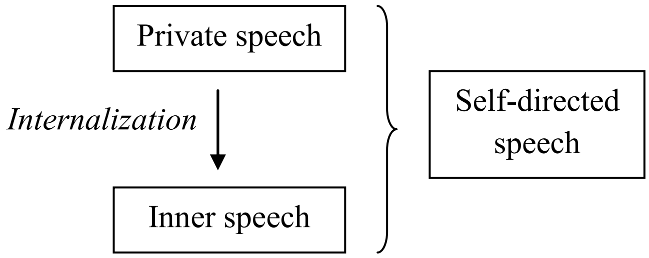

Figure 1. Conceptual relations between private speech, inner speech, and self-directed speech.

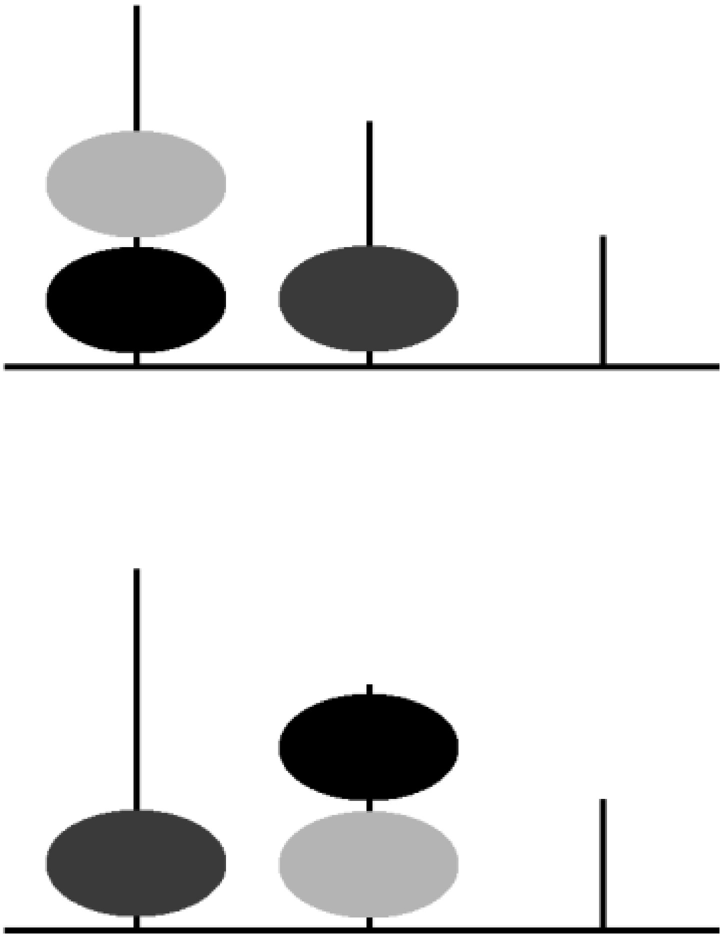

Figure 1. Example Tower of London problem. Top: start state. Bottom: goal state. Actual colors were red, green, and blue.

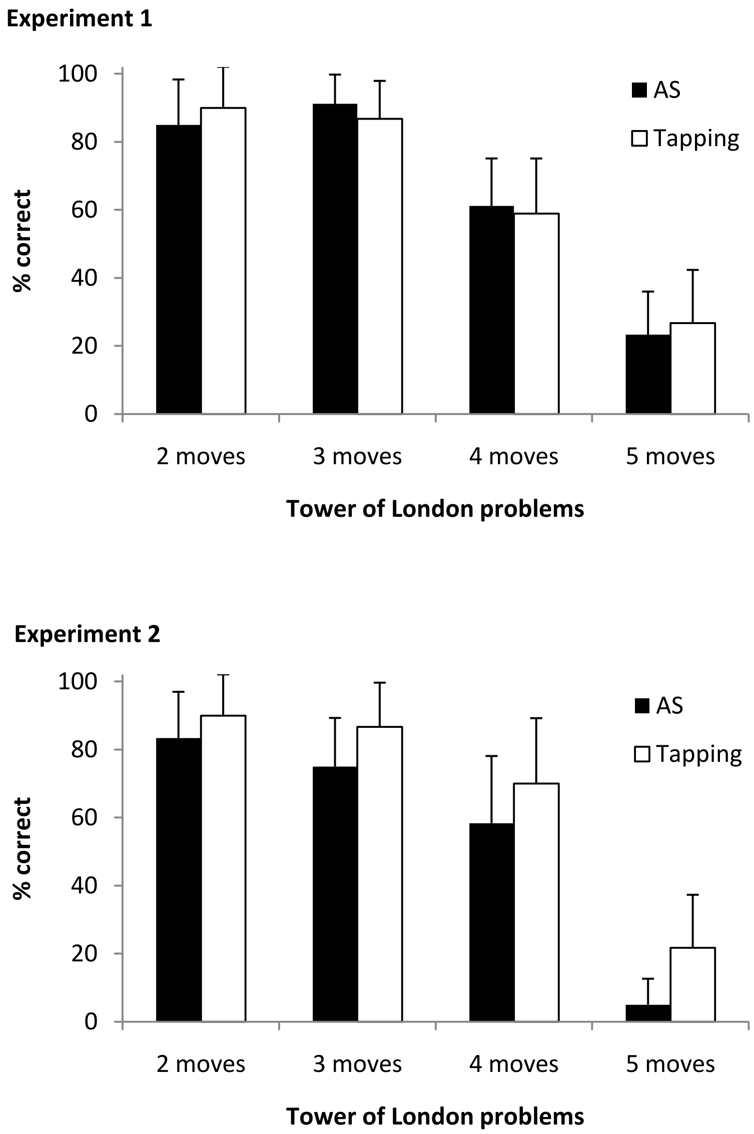

Figure 3. Percentage of Tower of London trials solved correctly, in the articulatory suppression (AS) and tapping conditions. Error bars indicate 0.5 × *SD*.

References

Al-Namlah, A. S., Fernyhough, C., & Meins, E. (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi Arabian and British samples. *Developmental Psychology, 42,* 117-131.

Ang, S. Y., & Lee, K. (2008). Central executive involvement in children's spatial memory. *Memory, 16*(8), 918-933.

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology-General, 130*(4), 641-657.

Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R., J., Frackowiak, R. S. J., & Robbins, T. W. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia, 34*(6), 515-526.

Baldo, J. V., Dronkers, N. F., Wilkins, D., Lucy., C., Raskin, P., & Kim, J. Y. (2005). Is problem solving dependent on language? *Brain and Language, 92*(3), 240-250.

Behrend, D. A., Rosengren, K., & Perlmutter, M. (1989). A new look at children's private speech: The effects of age, task difficulty, and parent presence. *International Journal of Behavioral Development, 12*(3), 305-320.

Behrend, D. A., Rosengren, K. S., & Perlmutter, M. (1992). The relation between private speech and parental interactive style. In R. M. Diaz & L. E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 85-100). Hillsdale, N.J.: Erlbaum.

Berg, W. K., & Byrd, D. L. (2002). The Tower of London spatial problem-solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology, 24*(5), 586-604.

Berk, L. E. (1986). Relationship of elementary school children's private speech to behavioral accompaniment to task, attention, and task performance. *Developmental Psychology, 22,* 671-680.

Boghi, A., Rasetti, R., Avidano, F., Manzone, C., Orsi, L., D'Agata, F., Caroppo, P., Bergui, M., Rocca, P., Pulvirenti, L., Bradac, G. B., Bogetto, F., Mutani, R., & Mortara, P. (2006). The effect of gender on planning: An fMRI study using the Tower of London task. *Neuroimage, 33*(3), 999-1010.

Cohen, G. (1996). *Memory in the real world.* Hove, UK: Psychology Press.

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language, 48*(1), 148-168.

Fernyhough, C. (2008). Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding. *Developmental Review, 28*(2), 225-262.

Fernyhough, C., & Fradley, E. (2005). Private speech on an executive task: relations with task difficulty and task performance. *Cognitive Development, 20*(1), 103-120.

Frauenglass, M. H., & Diaz, R. M. (1985). Self-regulatory functions of children's private speech: A critical analysis of recent challenges to Vygotsky's theory. *Developmental Psychology, 21*(2), 357-364.

Hermer-Vasquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology, 39*(1), 3-36.

Kaller, C. P., Unterrainer, J. M., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: Insights from the Tower of London task. *Cognitive Brain Research, 20,* 462-472.

Kim, H. S. (2002). We talk, therefore we think? A cultural analysis of the effect of talking on thinking. *Journal of Personality and Social Psychology, 83*(4), 828-842.

Lidstone, J. S. M., Fernyhough, C., Meins, E., & Whitehouse, A. J. O. (2009). Brief report: Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills. *Journal of Autism and Developmental Disorders, 39*, 1222-1225.

Liefooghe, B., Vandierendonck, A., Muyllaert, I., Verbruggen, F., & Vanneste, W. (2005). The phonological loop in task alternation and task repetition. *Memory, 13*(5), 550-560.

Owen, A. M., Sahakian, B. J., Hodges, J. R., Summers, B. A., Polkey, C. E., & Robbins, T. W. (1995). Dopamine-dependent fronto-striatal planning deficits in early Parkinson's disease. *Neuropsychology, 9,* 126–140.

Nakabayashi, K., & Burton, A. M. (2008). The role of verbal processing at different stages of recognition memory for faces. *European Journal of Cognitive Psychology, 20*(3), 478-496.

Phillips, L. H., Wynn, V. E., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the Tower of London task. *Memory, 7*(2), 209-231.

Phillips, L. H., Wynn, V. E., McPherson, S., & Gilhooly, K. J. (2001). Mental planning and the Tower of London task. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 54*(2), 579-597.

Robbins, T. W., Anderson, E. J., Barker, D. R., Bradley, A. C., Fernyhough, C., Henson, R., Hudson, S. R., & Baddeley, A. (1996). Working memory in chess. *Memory and Cognition, 24,* 83-93.

Saeki, E., Saito, S., & Kawaguchi, J. (2006). Effects of response-stimulus interval manipulation and articulatory suppression on task switching. *Memory, 14*(8), 965-976.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, Biology, 298,* 199-209.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* Boston: Allyn and Bacon.

Vygotsky, L. S. (1978). In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press (Original work published 1930, 1933, and 1935).

Vygotsky, L. S. (1987). Thinking and speech. In *The collected works of L. S. Vygotsky (Vol. 1)*. New York: Plenum (Original work published 1934).

Wallace, G. L., Silvers, J. A., Martin, A., & Kenworthy, L. E. (2009). Brief report: Further evidence for inner speech deficits in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(12), 1735-1739.

Ward, G., & Allport, A. (1997). Planning and problem-solving using the five-disc tower of London task. *Quarterly Journal of Experimental Psychology Section A – Human Experimental Psychology, 50*(1), 49-78.

Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry, 47*(8), 857-865.

Winsler, A. (2009). Still talking to ourselves after all these years: A review of current research on private speech. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 3-41). New York: Cambridge University Press.

Winsler, A., Diaz, R. M., & Montero, I. (1997). The role of private speech in the transition from collaborative to independent task performance in young children. *Early Childhood Research Quarterly, 12*(1), 59-79.

Winsler, A., Fernyhough, C., McClaren, E. M., & Way, E. (2005). *Private speech coding manual*. Unpublished manuscript. George Mason University, USA. Available at: http://classweb.gmu.edu/awinsler/PSCodingManual.pdf