Cognitive Development 26 (2011) 203–213

Contents lists available at ScienceDirect

Cognitive Development

Individual differences in children's private speech: Consistency across tasks, timepoints, and contexts

Jane Lidstone*, Elizabeth Meins, Charles Fernyhough

Durham University, United Kingdom

ARTICLE INFO

Keywords:
Private speech
Verbal mediation
Problem-solving
Cognitive development

ABSTRACT

Children often talk themselves through their activities. They produce private speech (PS), which is internalized to form inner speech (silent verbal thought). Twenty-five 8–10-year-olds completed four tasks in a laboratory context (Tower of London, digit span, and two measures of spatial IQ). PS production was recorded. Eleven months later, the same participants completed the Tower of London and academic numeracy tasks, again in a laboratory context, as well as numeracy tasks in a classroom context. Rates of PS production and its level of internalization showed large positive correlations across time, tasks, and contexts. The results are interpreted in terms of the psychometric properties of PS production and are taken as evidence for the development of a domain-general system for verbal self-regulation in childhood.

Private speech (PS) is the self-directed speech that emerges during the preschool years, when children start to talk themselves through their activities. Private speech appears to be functionally related to cognitive performance: It appears at times of difficulty with a task (Behrend, Rosengren, & Perlmutter, 1989; Duncan & Pratt, 1997; Fernyhough & Fradley, 2005; Kohlberg, Yaeger, & Hjertholm, 1968), and its production is associated with success on a variety of tasks (Behrend, Rosengren, & Perlmutter, 1992; Bivens & Berk, 1990; Fernyhough & Fradley, 2005; Winsler, Diaz, & Montero, 1997). The incidence of fully overt PS drops off during early childhood, and, in middle childhood, PS is more likely to take the form of covert muttering and whispering, and silent, verbal lip movements (Kohlberg et al., 1968; Winsler, 2009). This shift toward covert PS is thought to reflect the gradual internalization of PS to form inner speech, or verbal thought (Vygotsky, 1934/1987). As such, covert PS is considered more mature than fully overt PS. In the most widely used scheme for categorizing PS (Berk, 1986), the

* Corresponding author at: School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, CF10 3AT, UK.
E-mail address: lidstonej@cardiff.ac.uk (J. Lidstone).

0885-2014/$ – see front matter
doi:10.1016/j.cogdev.2011.02.002

least mature is task-irrelevant PS (Level 1), followed by task-relevant overt PS (Level 2), then partially internalized PS (Level 3).

1. A domain-general shift to verbal mediation

Vygotsky (1930–1935/1978) interpreted the emergence of PS as marking a radical reorganization of children's cognition:

The most significant moment in the course of intellectual development, which gives birth to the purely human forms of practical activity and abstract intelligence, occurs when speech and practical activity converge (p. 24).

He compares this transformation to that occurring in practical activity upon the introduction of tools. That cognition becomes subject to the organizing function of language, he argues, will be apparent in the domains of perception, attention, "thinking", and "active remembering"—in sum, most goal-directed cognitive activity. Therefore Vygotsky's claim was not merely that some activities become amenable to verbally mediated strategies but rather that, by middle childhood, goal-directed cognition is quite fundamentally verbal in nature. In today's terms we would say that Vygotsky predicted domain-general development of verbal mediation (Al-Namlah, Fernyhough, & Meins, 2006). Children's production of PS has been documented during a wide range of cognitive tasks, including problem-solving tasks (Behrend et al., 1992; Berk & Spuhl, 1995; Daugherty, White, & Manning, 1994; Winsler, de León, Wallace, Carlton, & Willson-Quayle, 2003; Winsler, Diaz, McCarthy, Atencio, & Adams Chabay, 1999), executive function tasks (Fernyhough & Fradley, 2005; Müller, Zelazo, Hood, Leone, & Rohrer, 2004; Winsler, Abar, Feder, Schunn, & Rubio, 2007; Winsler, Diaz, Atencio, McCarthy, & Adams Chabay, 2000; Winsler et al., 1997; Winsler & Naglieri, 2003; Winsler, Naglieri, & Manfra, 2006), and schoolwork in both language (Berk & Landau, 1993) and mathematics (Berk, 1986; Berk & Landau, 1993; Berk & Potts, 1991; Bivens & Berk, 1990; Ostad & Sorensen, 2007). That PS appears in such a broad range of tasks supports neoVygotskian ideas about the domain-generality of verbal mediation.

However, stronger evidence would come from an analysis of the extent to which individual differences in children's PS production are stable across different types of task. Cross-task correlations would imply that PS represents "not just moment-to-moment articulation of ongoing thought processes during task-specific problem solving, but instead a coherent set of verbal self-regulatory strategies that have developed over time into an organized way of guiding one's behavior" (Winsler, 2009, p. 8).

Cross-task correlations in PS production have, to our knowledge, been reported in only one published study. Winsler et al. (2003) investigated the consistency of individual differences in 32 preschoolers' PS production across two tasks. The first was a selective attention task; for each trial participants viewed two pictures and indicated, by choosing a third picture, the attribute the other two pictures shared. In the second task, participants attempted to reproduce a Lego structure according to an accessible three-dimensional model previously constructed in collaboration with the experimenter. The authors reported a correlation of .70 between the two tasks in the rate of PS production (utterances per minute). The proportion of PS of each developmental level showed some limited cross-task consistency – correlations for Levels 1, 2 and 3 of .02, .33, and .62 respectively. Only the last of these was statistically significant.

At present there thus exists some limited evidence of domain-generality of PS. One surprisingly neglected aspect of PS is short-term memory. Some memory tasks such as remembering words and digits are by definition verbally mediated. However, very little is known about children's PS production during such tasks. Our knowledge is limited to the observation that young children produce PS when presented with speech-based information (e.g., color names) to remember (Patrick & Abravanel, 2000; Winsler, Manfra, & Diaz, 2007). Whether their doing so is related to the production of PS during other tasks is not known.

2. Psychometric properties of private speech

The cross-task consistency of PS takes on particular importance from a methodological perspective. In much PS research, a single measurement of children's PS production is taken as representative of their PS production in general. This is particularly the case in studies examining the developmental significance of PS production, such as those relating PS production to individual differences in measures of self-regulation or theory of mind, and studies comparing PS production of children with developmental disorders to that of typically developing children. Researchers assume that, had PS been recorded during a slightly different task, individual differences in PS production would have been similar (Winsler, 2009; Winsler et al., 2003). Similarly, it is assumed that measuring PS production at a different time, or in a context other than the laboratory, would produce similar results. These key issues relating to the psychometric properties of PS production have been largely neglected to date. To the best of our knowledge there exist only two published studies, Winsler et al.'s (2003) study described above, which also addressed longitudinal stability, and a study by Berk and Landau (1993), which examined cross-context consistency.

Winsler et al. (2003) addressed the cross-timepoint stability of individual differences in PS by having preschoolers complete selective attention and Lego construction tasks on a second occasion, six months after the first session. Individual differences in the rate of PS (in utterances per minute) were preserved across time ($r = .35$), although the correlation was smaller than the cross-task correlation noted above. The cross-timepoint correlations for PS of Levels 1, 2, and 3 were .07, .39, and .28, respectively. Only the largest of these was statistically significant. There is therefore some equivocal support for the cross-timepoint consistency of individual differences in PS production.

Although a degree of longitudinal stability in PS measurements is expected, one would also predict individual differences in the rate of development. That is, measures of PS might increase for the group as a whole (Winsler et al., 2003), but the rate of change might be greater for some children than others. We imagine this to be particularly true in early childhood, when there is rapid change in PS. In middle childhood, when the shift to verbal mediation has been accomplished, we might expect greater cross-timepoint consistency.

The only evidence relating to the consistency of individual differences in PS across contexts comes from a study of 14 normally achieving children and 14 learning disabled children, all 9–12 years old (Berk & Landau, 1993). They were observed while engaged in academic (language and math) tasks in the classroom and in the laboratory. The authors report a significant cross-context correlation ($r = .58$) for the rate of Level 3 PS production (measured in terms of the proportion of 10-s periods containing PS). However, PS of Levels 1 and 2 was not reported as it was relatively rare, so it is difficult to tell whether or not individual differences in the overall rate of PS and its internalization level remained consistent across contexts.

In sum, there is very little evidence regarding the psychometric properties of PS, and what exists is somewhat inconclusive. Given the importance of these questions for both theory and method in this area, the principal aim of the present study was to provide further evidence regarding the consistency of individual differences in PS across tasks, timepoints, and contexts, in a sample of typically developing children.

2.1. Present study

At an initial session, 8–10-year-olds completed four tasks: the Tower of London, a planning task known to be verbally mediated in middle childhood (Lidstone, Meins, & Fernyhough, 2010); digit span, a short-term memory task known to elicit phonological (verbal) rehearsal in middle childhood; and two tests of spatial IQ, one involving short-term memory for abstract line drawings, and the other similar to the Block Design subtest of the Wechsler Intelligence Scales. These tasks were chosen to draw on a range of cognitive functions, providing a test of the domain-generality of PS development. At the second session, participants completed the Tower of London task a second time and some numeracy school work. All tasks were completed in a laboratory context, except for half of the numeracy work, which was completed in the classroom.

We hypothesized that individual differences in the rate of PS production and its internalization level would (a) be consistent across tasks (within timepoints, in the laboratory context), (b) remain stable over time (on the task completed twice), and (c) be consistent across contexts (classroom versus laboratory contexts).

In Berk and Landau's study of PS in middle childhood, much of the PS fell into the "partially internalized" category, preventing the authors from investigating the extent to which individual differences in the internalization level of PS were preserved across contexts. In other words, coding of task-relevant PS as simply overt or partially internalized did not allow individual differences in internalization level to be distinguished. We therefore developed a more fine-grained scale to measure the internalization level of PS, anticipating variation within as well as between categories, particularly the Level 3 category which encompasses muttering, whispering, and completely silent lip movements. We therefore endeavoured to make use of these distinctions implicit in the wording of Berk's (1986) coding scheme. Subsidiary aims of this study were therefore to develop a sensitive ordinal scale measuring the internalization level of PS, and to establish its inter-rater reliability and convergent validity. Convergent validity was assessed by examining correlation between internalization scores and chronological age; PS should become more internalized with increasing age.

3. Method

3.1. Participants

The participants were 25 typically developing 8–10-year-olds, recruited from three mainstream state schools in the North-East of England. Two of the schools were in moderately disadvantaged areas and the other was in a moderately affluent area (www.ofsted.gov.uk). The initial assessment was conducted in conjunction with two other studies on the verbal mediation of Tower of London performance. At the second assessment a mean of 11 months later (range 9–12 months), two of an original group of 30 had moved away from the area and three declined to participate in the present study. Our sample therefore consisted of 25 participants (13 girls). At Time 1, their mean age was 9–4 years (range 8–0 to 10–9).

3.2. Design

At Time 1, participants completed eight Tower of London problems, a digit span task, and two subtests of the British Ability Scales: Recall of Designs, and Pattern Construction (Elliott, Smith, & McCullough, 1996). The tasks were completed in this fixed order in two sessions conducted within about two weeks of each other. All Time 1 tasks were completed in a laboratory context (see below for details). At Time 2, participants completed 12 Tower of London problems, and 20 minutes' worth of whatever numeracy schoolwork was scheduled for the day of data collection. The Tower of London problems and half of the numeracy work were completed in a laboratory context; the other half of the numeracy work was completed in the classroom (see below for details). The three observations at Time 2 (Tower of London in the laboratory context, numeracy work in the laboratory context, and numeracy work in the classroom) were completed in three separate sessions. The tasks were not completed in a fixed order at Time 2, as we had to fit observation of numeracy work around teachers' timetables, but the tasks at Time 2 were completed within a period of two weeks.

3.3. Tasks

Tower of London. Participants performed the three-disk Tower of London (Shallice, 1982), completing eight 2–5-move problems at Time 1 and twelve 3- to 5-move problems at Time 2. The problem set was more difficult at Time 2 in anticipation of children's increased proficiency at Time 2. No problem appeared in a problem set twice, and none of the problems presented at Time 2 duplicated those presented at Time 1. The standard Tower of London procedure requires participants to move the disks one at a time to make two configurations match (Shallice, 1982). The present study used a modified version of the task designed to encourage participants to make full mental plans (Baker et al., 1996; Boghi et al., 2006; Owen et al., 1995), as previous work indicated that children in this age group spend little time planning unless forced to do so by being prevented from moving the disks (Lidstone et al., 2010). Instead of asking participants to move the disks to make the configurations match (as per Shallice, 1982), we therefore asked them to plan the moves mentally. After telling the experimenter the number of moves they had planned, the participants were asked to demonstrate the moves (Lidstone et al., 2010). The purpose of this performance phase was to verify that participants had actually planned the moves rather than simply guessing a number. Only the time between presentation of the problem and participants' verbal numerical response was of interest in terms of PS, as this is when planning took place. Therefore only the planning phase was coded for PS.

Digit span. Participants completed the digit span task once in each of the two conditions as part of a dual-task paradigm for another study. In one condition they tapped their foot while doing the task. In the other they repeated the word "Monday" instead of foot-tapping. The order of the conditions was counterbalanced. Only the results from the tapping condition are used in the present study. The digit span task was based on that of Chincotta and Chincotta (1996). Digits were presented on a laptop computer screen at a rate of one per second. After the last digit, there was a blank screen for 4 s, and then a question mark appeared, at which point the participant was required to recall the digits orally in the order in which they had been presented. Only the period between the start of the trial and the presentation of the question mark was coded for PS. The trials were organized in blocks of three sequences of the same length, starting with sequences of two digits. Participants proceeded to the next level if and when they had recalled two sequences of the current length correctly.

Recall of Designs. For each trial, participants viewed an abstract line drawing for 5 s and then attempted to reproduce it from memory on squared paper. Participants started at Trial 1 or 3 depending on their age (as per the British Ability Scales manual), and continued until Trial 14 or until they scored zero on five consecutive trials.

Pattern Construction. For each trial, participants were required to assemble flat squares or solid cubes to create a larger two-dimensional pattern to match a picture, which remained in view until the end of the trial. On the first seven trials, between two and six squares, yellow on one side and black on the other, were presented. On subsequent trials, there were between two and nine six-sided blocks, each block having four different sides of all yellow, all black, or a combination. Participants started at Trial 1 or 8, depending on their age, and continued to Trial 20 or until they failed to create the required pattern within the time limit on five consecutive trials. The period coded for PS ended either when the participant indicated they had finished or when the time limit was reached.

Numeracy. Participants engaged in whatever numeracy work was scheduled for that day at school. The teacher typically started the session by tutoring the whole class on the topic. Then class members individually practised the skills learned. Tasks included practising written methods of addition, subtraction, multiplication, and division; deducing, measuring, and drawing angles; and drawing and interpreting tables and graphs. Half of this work was done by participants in the classroom as normal, the other half in the laboratory context. Often, there were only 5 min of individual work per numeracy lesson and so observation occurred over several days.

3.4. *Observation*

To achieve the unfamiliarity of a laboratory context, participants worked in rooms of the schools that the children were not normally permitted to enter. They completed the tasks individually with the experimenter seated at the same table, providing general encouragement at intervals. A camcorder recorded PS.

In the classroom, a webcam was used as it was smaller than the camcorder and could be securely attached to participants' desks. Participants were aware they were being filmed at all times but otherwise worked in a normal classroom setting. This setting differed from the laboratory context in three respects – the physical environment was very familiar, participants worked in parallel with (and within earshot of) their peers, and there was no immediate adult presence. Participants were seated at Tables of 2–6 children. The children were generally required to work on their own but they were permitted to talk quietly to each other "if they got stuck." Participants did not appear to find the camera distracting or inhibiting in either context.

3.5. *Coding*

PS was coded from the video recordings. On the basis of Berk and Landau's (1993) results, we anticipated negligible amounts of task-irrelevant PS, so PS was defined as speech, including muttering,

Table 1
Number of observation periods for each task.

Task	Mean	*SD*	Range
Time 1			
Tower of London	8	0	–
Digit span	9	2	6–12
Recall of Designs	10	2	6–12
Pattern Construction	13	2	9–16
Time 2			
Tower of London	20	6	12–39
Numeracy: laboratory context	66	25	21–127
Numeracy: classroom	60	24	16–114

Note. All tasks were completed in a laboratory context, apart from one of the numeracy sessions, which was completed in the classroom. At Time 1, an observation period was a trial. At Time 2, an observation period was 10 s of on-task time.

whispering, and verbal silent lip movements, that was relevant to the task, and not directed toward the experimenter (in the laboratory context) or peers or teachers (in the classroom).

To calculate the rate of PS production, on-task time was divided into observation periods. At Time 1, an observation period was simply a trial. Numeracy work was not divided into trials so, at Time 2, an observation period was defined as a period of on-task time lasting 10 s (after Berk & Landau, 1993, and other studies measuring children PS production during completion of school work; Berk, 1986, Berk & Potts, 1991). Berk and Landau employed live observation; hence of every 30 s, the first 10 were spent observing, and the next 20 recording PS codes. As we had video recordings, we included three 10-s observation periods in every 30 s of on-task time. Similarly to Berk and Landau, off-task time included (a) time spent watching, listening to, or interacting with peers and teachers and, in the laboratory context, the experimenter, and (b) during numeracy work, time spent sharpening pencils, finding erasers, etc. The mean number of observation periods for each task is shown in Table 1.

The rate of PS production was quantified as the percentage of observation periods that contained PS. To investigate the effect of mode of defining an observation period (trial vs. time period), we quantified the rate of PS production on the Tower of London at Time 2 using both definitions. The correlation between them was very large, $\rho(25) = .97$, $p < .001$, and the pattern of correlations with other variables was exactly the same.

Where PS was present during an observation period, it was coded in terms of its level of internalization. PS is traditionally (Winsler, Fernyhough, McClaren, & Way, 2005) coded according to Berk's (1986) three-level scheme, as Level 1 (task-irrelevant private speech), Level 2 (task-relevant, overt private speech), or Level 3 (external manifestations of task-relevant inner speech, including inaudible muttering and whispering, and silent, verbal lip movements). As there were only two task-irrelevant utterances in the present corpus, these were excluded from analysis. Some examples of the PS produced are:

One, two, three, no, one . . .
That one's going there
This one's in the way
This isn't right
How many moves.

The categories of task-relevant speech implicit in Berk's (1986) coding scheme were defined as follows. Muttering and overt speech were defined as speech that is audible because it is voiced. *Muttering* could be intelligible or unintelligible. Intelligible muttering was distinguished from overt speech as significantly quieter and/or more indistinct than the child's social speech between trials. *Whispering* was defined as unvoiced speech, audible not because it is voiced but because of the adduction of vocal cords produced by the exhalation of breath. Silent verbal *lip movements* were defined as lip movements where there was clear evidence that words were being uttered. In practice it was difficult to distinguish totally silent lip movements from those accompanied by small sounds produced by the

Table 2
Derivation of five-level internalization coding scheme from three dimensions of covertness.

Level	Volume[a]	Intelligible[b]	Voiced[b]	Total score	Description
1	0	0	0	0	Fully overt speech
2	1	0	0	1	Intelligible muttering
3	1	1	0	2	Intelligible whispering
	1	0	1	2	OR unintelligible muttering[c]
4	1	1	1	3	Audible but unintelligible whispering
5	2	1	1	4	Inaudible and barely audible verbal lip and tongue movements

Notes. Not all theoretically possible combinations of the three dimensions are represented here because not all combinations are practically possible: e.g., if an utterance is voiced it must be audible.

[a] Three levels: 0 (full volume, the same as social speech); 1 (quieter than social speech); 2 (inaudible and barely audible, which were difficult to distinguish from each other therefore combined).

[b] Two levels: 0 (yes); 1 (no).

[c] There was no rationale for categorizing either of these types of utterance as more internalized than the other.

interaction of mouth parts. This category was therefore redefined as inaudible and barely audible verbal lip movements. Barely audible lip movements are distinguishable from whispering because, in the latter, vowel sounds are audible.

Three dimensions of covertness were extracted from these definitions: (a) volume of speech, (b) whether or not speech was voiced, and (c) whether or not speech was intelligible. When combined (Table 2), these dimensions yield five levels of internalization:

Level 1	Fully overt speech
Level 2	Intelligible muttering
Level 3	Intelligible whispering OR unintelligible muttering
Level 4	Audible but unintelligible whispering
Level 5	Inaudible and barely audible verbal lip and tongue movements

If participants spoke at all during a task, they were given an internalization score for that task as follows. Each observation period including PS was given an internalization score. The internalization score for an observation period was the mean of all the speech that occurred during that observation period. For example, a 10-s period containing 3 s of Level 2 speech, 3 s of Level 3 speech, and 4 s of no speech, was scored 2.5, as half the speech was of Level 2 and half the speech of Level 3. Observation periods with no speech were not given internalization scores. A participant's internalization score for a task was the mean score for all the observation periods including PS. The range of possible internalization scores was thus 1.0–5.0, with higher scores indicating more internalized PS.

Because PS was relatively rare during the Recall of Designs and Pattern Construction tasks (PS was produced by fewer than five participants), internalization scores are not reported for these tasks. For the other tasks, analyses of internalization scores are restricted to participants who produced some PS during the tasks in question. The sound quality of three of the recordings of numeracy work in the classroom was not sufficient to allow coding of internalization levels, but the coding of PS rates was unaffected.

A second researcher independently coded 20% of the recordings, five for each task, to assess inter-rater reliability. For the presence/absence of PS during an observation period, the coefficient of agreement (Cohen's κ) was .86. The Spearman correlation between the two coders' internalization scores was $\rho(254) = .89$, $p < .001$.

4. Results

4.1. Preliminary analyses

The distribution of nine of the 12 variables differed from normal (Shapiro–Wilk tests, $p < .05$), so all statistical tests are nonparametric. All hypotheses are tested with Spearman's rank correlation

Table 3
Rate of private speech production and internalization level of private speech: descriptive statistics.

Task	Rate of private speech production[a]			Internalization level of private speech			
	Mean	*SD*	Range	*N*	Mean	*SD*	Range
Time 1							
Tower of London	38.5	31.4	0.0–100.0	19	4.8	0.5	3.7–5.0
Digit span	50.8	32.5	0.0–100.0	21	4.9	0.4	3.5–5.0
Recall of Designs	19.6	32.4	0.0–100.0	–	–	–	–
Pattern Construction	18.7	29.3	0.0–94.0	–	–	–	–
Time 2							
Tower of London	63.0	39.0	0.0–100.0	21	4.3	0.9	2.4–5.0
Numeracy: laboratory context	45.0	35.3	0.0–100.0	24	3.9	1.0	2.2–5.0
Numeracy: classroom	48.4	23.5	12.1–97.6	22	3.2	1.0	1.5–4.8

Note. $N = 25$ unless otherwise shown.
[a] Percentage of observation periods with private speech.

coefficient (ρ). In the preliminary analyses, related samples are compared with the Wilcoxon signed-rank test (Z). All tests are two-tailed.

The mean rate of PS production varied between 18.7% and 63.0% across tasks (Table 3). The range of PS rates was large for all tasks.

The rate of PS production was compared across tasks, across timepoints, and across contexts, with α adjusted for multiple comparisons ($\alpha = .05/7 = .007$). Cross-task comparisons within timepoints and contexts showed that PS was more frequent during the tasks known to be verbally mediated (Tower of London, digit span) than on the two spatial IQ tasks, $Zs > 3.70$, $ps < .001$. No other cross-task comparisons within timepoints and contexts showed significant differences, $Zs < 2.21$, ns. The cross-timepoint comparison showed that the rate of PS production on the Tower of London was higher at Time 2 than at Time 1, $Z = 4.11$, $p < .001$. The rate of PS production did not vary across contexts for numeracy work, $Z = 0.31$, $p = .76$.

Mean internalization scores at Time 1 were very high, almost exclusively Level 5 PS. On the digit span task, all but four participants had internalization scores of 5.0, yielding insufficient variation for correlation analyses. Therefore only Time 2 internalization scores were used to determine cross-task stability of the internalization level of PS.

Negative correlations provide evidence for convergent validity of the internalization scale, although not all coefficients were statistically significant. For the Tower of London at Time 2, $\rho(21) = -.39$, $p = .08$; for numeracy in the laboratory context, $\rho(24) = -.40$, $p = .05$; for numeracy in the classroom, $\rho(22) = -.29$, $p = .19$.

Internalization scores were compared across tasks, across timepoints, and across contexts, with α adjusted for multiple comparisons ($\alpha = .05/4 = .0125$). The PS produced during Tower of London performance at Time 2 was more internalized than that produced in numeracy work in the laboratory context, $Z = 2.68$, $p = .007$, which was more internalized than that produced during numeracy work in the classroom, $Z = 2.99$, $p = .003$. There were no other cross-task or cross-timepoint differences in internalization levels, $Zs < 1.94$, ns.

4.2. Consistency of individual differences across tasks, timepoints, and contexts

In terms of the rate of PS production, there were large positive correlations among all tasks at Time 1 (Table 4). At Time 2, there was a positive correlation between the rate of PS production during numeracy work completed in the laboratory context and the rate of PS production during Tower of London performance, $\rho(25) = .51$, $p = .01$. With regard to internalization levels, there was a positive correlation between Time 2 scores on the Tower of London and scores during the numeracy tasks in the laboratory context, $\rho(21) = .47$, $p = .03$.

Comparison of individual differences across timepoints revealed a positive relation between the rate of PS production on the Tower of London at Time 1 and that during Tower of London performance

Table 4
Cross-task correlations among private speech rates during tasks completed at Time 1.

	ToL	DS	RoD	PC
Tower of London (ToL)	–	–	–	–
Digit span (DS)	.59**	–	–	–
Recall of Designs (RoD)	.71***	.55**	–	–
Pattern Construction (PC)	.64**	.47*	.65***	–

Note. $N = 25$. Private speech rate is the percentage of observation periods containing private speech.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

at Time 2, $\rho(25) = .54$, $p = .01$. In addition there was a positive correlation between internalization scores on the Tower of London at Time 1 and the Tower of London at Time 2, $\rho(21) = .57$, $p = .02$.

Turning to consistency across contexts, there was a positive correlation between the rate of PS production during numeracy work in the laboratory context and that during numeracy work in the classroom, $\rho(25) = .53$, $p = .01$. Furthermore, there was a positive correlation between internalization scores on numeracy in the laboratory context and numeracy in the classroom, $\rho(22) = .58$, $p = .01$.

5. Discussion

The rate of PS production and its internalization level showed strong consistency across tasks, timepoints, and contexts. The present findings on cross-task consistency in PS in middle childhood thus accord with those of Winsler et al.'s (2003) study of preschoolers. The results indicate that measures obtained from a single task can be a reliable guide to a child's PS relative to that of peers. Although group rates of PS production varied by task (e.g., PS was produced more often during the Tower of London than the Spatial IQ tasks), children's scores relative to one another were preserved.

The present findings are the first relating PS production on a memory task (digit span) to PS production on problem-solving and executive-function tasks (but see also Al-Namlah et al., 2006). The correlations across tasks can be viewed as supporting the claim that development of verbal mediation is domain general (Al-Namlah et al., 2006). Assuming that more highly internalized PS indicates more advanced verbal mediation, children who showed more advanced PS use on one task showed similarly advanced PS use on tasks in different domains. Alternatively, we could view lower rates of PS use as indexing more advanced verbal mediation, as, in the age range studied here, children are usually considered to be past the peak of private speech production (Fernyhough & Meins, 2009). In either case, the findings that both production and internalization scores were correlated across tasks support the idea of a domain-general change with respect to verbal mediation.

The present study showed strong evidence of consistency of individual differences over 11 months during middle childhood, a finding in line with Winsler et al.'s (2003) findings for preschoolers across six months. Winsler et al. found a cross-timepoint correlation in PS rate (in utterances per minute) of .35, and cross-timepoint correlations in proportion of PS of Levels 2 and 3 of .39 and .28, respectively, compared to our correlations in both rate and internalization level exceeding .50. These correlations occur despite group-level changes in quantity and characteristics of PS over time.

The finding of cross-context consistency of PS production on academic tasks replicated those of Berk and Landau (1993). Our more fine-grained analysis of internalization level allows us also to conclude that individual differences in internalization level remain consistent across contexts. Thus, it can be concluded that asking children of this age to do an appropriate task in a slightly "artificial" context does not alter PS.

In sum, our findings indicate that rate and internalization level of a child's PS in performing a task can be representative of the PS they would produce on another task, at another timepoint, or in another context. A subsidiary aim of the present work was to develop a new scale for measuring the internalization level of PS. Results indicated good inter-rater reliability and convergent validity in the form of negative correlations with age. In future work, to further validate the internalization scale, researchers could examine age trends in a longitudinal design. A challenge would be to keep the difficulty level of the tasks constant relative to the child's developmental level. An alternative approach would be to examine trends on a more microdevelopmental timescale.

A potential shortcoming of the internalization scale is that it produced skewed distributions not amenable to parametric analyses. Negligible variation in internalization scores on the digit span task precluded analysis of these data. It would be worthwhile to see if this age group produces such highly internalized speech as measured by this scale in future studies. On other tasks, however, the scale was effective in measuring individual differences in PS internalization level. Future work could also explore the utility of the internalization scale in studying PS in larger samples and in other age groups, particularly younger children's presumably less internalized PS. At present we are limited in our ability to generalize our findings beyond middle childhood.

Other findings of interest include the effect of context on PS internalization levels, despite the consistency of individual differences across contexts. Children may feel more inhibited in the laboratory context. Or the noisier classroom context may mean that children's PS has to be more overt in the classroom to have the same effect. The fact that PS was more overt in the classroom speaks against the idea that the main factor driving internalization is the transition to school (Duncan & Tarulli, 2009). Children in the present study used PS freely during numeracy lessons in the classroom.

Although the spatial IQ tasks yielded lower PS rates than the tasks known to be verbally mediated, PS was found during 20% of the spatial IQ task trials on average, and it appeared to be meaningfully related to the PS produced during verbally mediated tasks. The idea that some spatial or nonverbal IQ tasks are verbally mediated is supported by a finding that suppression of verbal processes during adults' performance of Raven's matrices is deleterious to performance (Kim, 2002).

Finally, we return to the issue of domain-generality in the development of verbal mediation. We have argued that cross-task correlations suggest a domain-general system for verbal mediation. More broadly, neoVygotskian theory predicts PS to be a pervasive feature of children's cognition, evident in multiple domains. Such pervasiveness is supported by the present findings. The rate of PS production was relatively high for all tasks and did not diminish over the 11-month period, in line with previous research on the persistence of PS into middle childhood and beyond (Duncan & Tarulli, 2009). Together with studies indicating that PS is useful for task performance (Lidstone et al., 2010; Winsler, 2009), our findings underline the significance of PS in children's cognitive development.

In sum, the present study indicates that individual differences in both the rate of PS production and its internalization level remain consistent across tasks, timepoints, and contexts in middle childhood. Our conclusion, which needs to be further substantiated by continued research on this topic, is that the large cross-task correlations are consistent with the idea of the development of a domain-general system for the verbal mediation of cognition.

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.

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

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*, 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, NJ: Erlbaum.

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.

Berk, L. E., & Landau, S. (1993). Private speech of learning-disabled and normally achieving children in classroom academic and laboratory contexts. *Child Development*, *64*, 556–571.

Berk, L. E., & Potts, M. K. (1991). Development and functional significance of private speech among attention-deficit hyperactivity disordered and normal boys. *Journal of Abnormal Child Psychology*, *19*, 357–377.

Berk, L. E., & Spuhl, S. T. (1995). Maternal interaction, private speech, and task-performance in preschool children. *Early Childhood Research Quarterly*, *10*, 145–169.

Bivens, J. A., & Berk, L. E. (1990). A longitudinal study of the development of elementary school children's private speech. *Merrill-Palmer Quarterly*, *36*, 443–463.

Boghi, A., Rasetti, R., Avidano, F., Manzone, C., Orsi, L., D'Agata, F., et al. (2006). The effect of gender on planning: An fMRI study using the Tower of London task. *Neuroimage*, *33*, 999–1010.

Chincotta, M., & Chincotta, D. (1996). Digit span, articulatory suppression, and the deaf: A study of the Hong Kong Chinese. *American Annals of the Deaf*, *141*, 252–257.

Daugherty, M., White, C. S., & Manning, B. H. (1994). Relationships among private speech and creativity measurements of young children. *Gifted Child Quarterly*, *38*, 21–26.

Duncan, R., & Pratt, M. (1997). Microgenetic change in the quality and quantity of preschoolers' private speech. *International Journal of Behavioral Development*, *20*, 367–383.

Duncan, R., & Tarulli, D. (2009). On the persistence of private speech: Empirical and theoretical considerations. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 176–187). New York: Cambridge University Press.

Elliott, C. D., Smith, P., & McCullough, K. (1996). *British Ability Scales* (2nd ed.). London: nferNelson.

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

Fernyhough, C., & Meins, E. (2009). Private speech and theory of mind: Evidence for developing interfunctional relations. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 95–104). New York: Cambridge University Press.

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*, 828–842.

Kohlberg, L., Yaeger, J., & Hjertholm, E. (1968). Private speech: Four studies and a review of theories. *Child Development*, *39*, 691–736.

Lidstone, J. S. M., Meins, E., & Fernyhough, C. (2010). The roles of private speech and inner speech in planning in middle childhood: Evidence from a dual task paradigm. *Journal of Experimental Child Psychology*, *107*, 438–451.

Müller, U., Zelazo, P. D., Hood, S., Leone, T., & Rohrer, L. (2004). Interference control in a new rule use task: Age-related changes, labeling, and attention. *Child Development*, *75*, 1594–1609.

Ostad, S. A., & Sorensen, P. M. (2007). Private speech and strategy-use patterns: Bidirectional comparisons of children with and without mathematical difficulties in a developmental perspective. *Journal of Learning Disabilities*, *40*, 2–14.

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

Patrick, E., & Abravanel, E. (2000). The self-regulatory nature of preschool children's private speech in a naturalistic setting. *Applied Psycholinguistics*, *21*, 45–61.

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

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. The collected works of L.S. Vygotsky* New York: Plenum. (Original work published 1934).

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., Abar, B., Feder, M. A., Schunn, C. D., & Rubio, D. A. (2007). Private speech and executive functioning among high-functioning children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*, 1617–1635.

Winsler, A., De León, J. R., Wallace, B., Carlton, M. P., & Willson-Quayle, A. (2003). Private speech in preschool children: Developmental stability and change, across-task consistency, and relations with classroom behavior. *Journal of Child Language*, *30*, 583–608.

Winsler, A., Diaz, R. M., Atencio, D. J., McCarthy, E. M., & Adams Chabay, L. (2000). Verbal self-regulation over time in preschool children at-risk for attention and behavior problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 875–886.

Winsler, A., Diaz, R. M., McCarthy, E. M., Atencio, D., & Adams Chabay, L. (1999). Mother–child interaction, private speech, and task performance in preschool children with behavior problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*, 891–904.

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*, 59–79.

Winsler, A., Fernyhough, C., McClaren, E. M., & Way, E. (2005). Private speech coding manual. Unpublished manuscript. Department of Psychology, George Mason University, Fairfax, VA. Retrieved from http://classweb.gmu.edu/awinsler/PSCodingManual.pdf.

Winsler, A., Manfra, L., & Diaz, R. M. (2007). "Should I let them talk?" Private speech and task performance among preschool children with and without behavior problems. *Early Childhood Research Quarterly*, *22*, 215–231.

Winsler, A., & Naglieri, J. A. (2003). Overt and covert verbal problem-solving strategies: Developmental trends in use, awareness, and relations with task performance in children age 5 to 17. *Child Development*, *74*, 659–678.

Winsler, A., Naglieri, J. A., & Manfra, L. (2006). Children's search strategies and accompanying verbal and motor strategic behavior: Developmental trends and relations with task performance among children age 5 to 17. *Cognitive Development*, *21*, 232–248.