An Exploratory Study into Trade-off Effects of Complexity, Accuracy, and Fluency on Young Learners' Oral Task Repetition

Evelyn Sample & Marije Michel

Studying task repetition for adult and young foreign language learners of English (EFL) has received growing interest in recent literature within the task-based approach (Bygate, 2009; Hawkes, 2012; Mackey, Kanganas, & Oliver, 2007; Pinter, 2007b). Earlier work suggests that second language (L2) learners benefit from repeating the same or a slightly different task. Task repetition has been shown to enhance fluency and may also add to complexity or accuracy of production. However, few investigations have taken a closer look at the underlying relationships between the three dimensions of task performance: complexity, accuracy, and fluency (CAF). Using Skehan's (2009) trade-off hypothesis as an explanatory framework, our study aims to fill this gap by investigating interactions among CAF measures. We report on the repeated performances on an oral spot-the-difference task by six 9-year-old EFL learners. Mirroring earlier work, our data reveal significant increases of fluency through task repetition. Correlational analyses show that initial performances that benefit in one dimension come at the expense of another; by the third performance, however, trade-off effects disappear. Further qualitative explanations support our interpretation that with growing task-familiarity students are able to focus their attention on all three CAF dimensions simultaneously.

Au sein de la littérature relative à l'approche fondée sur les tâches, on évoque de plus en plus d'études portant sur la répétition des tâches pour l'enseignement de l'anglais langue étrangère aux jeunes et aux adultes (Bygate, 2009; Hawkes, 2012; Mackey, Kanganas, & Oliver, 2007; Pinter, 2007b). Des études antérieures semblent indiquer que les apprenants en L2 profitent de la répétition de la même tâche ou d'une tâche légèrement différente. Il a été démontré que la répétition des tâches améliore la fluidité et qu'elle pourrait augmenter la complexité ou la précision de la production. Toutefois, peu d'études se sont penchées davantage sur les relations sous-jacentes entre les trois dimensions de l'exécution des tâches : la complexité, la précision et la fluidité. S'appuyant sur l'hypothèse du compromis de Skehan (2009) comme cadre explicatif, notre étude vise à combler cette lacune en examinant les interactions entre les mesures de ces trois éléments. Nous faisons rapport du rendement de six jeunes âgés de 9 ans qui apprennent l'anglais comme langue étrangère alors qu'ils répètent une tâche impliquant l'identification de différences. Nos données reproduisent les résultats de travaux antérieurs en ce qu'elles révèlent une amélioration significative de la fluidité par la répétition

*de tâches. Des analyses corrélationnelles indiquent que l'amélioration d'une di-
mension lors des exécutions initiales se fait aux dépens d'une autre; cet effet de
compromis disparait, toutefois, à la troisième exécution. Des explications quali-
tatives supplémentaires viennent appuyer notre interprétation selon laquelle la
familiarité croissante que ressentent les élèves avec une tâche leur permet de se
concentrer sur les trois dimensions (complexité, précision et fluidité) à la fois.*

This exploratory study aims at broadening our understanding of how task-based language teaching (TBLT) pedagogy can help young learners[1] develop their language through task repetition. It explores the performance of 9-year-old second language (L2) learners on a series of spot-the-difference tasks.

Teaching English as a Foreign Language (EFL) to young learners "may be associated with a different set of challenges than those described in the more general education literature" (Stelma & Onat-Stelma, 2010, p. 195), such as finding the balance between encouragement, autonomy, and challenge on one hand and providing support and care on the other. Due in part to these specific needs, TBLT may be particularly appropriate for young learners (Bourke, 2006). The task-based approach promotes learning a language by means of tasks, that is, "a holistic activity, which engages language use in order to achieve some nonlinguistic outcome while meeting a linguistic challenge, with the overall aim of promoting language learning, through process or product or both" (Samuda & Bygate, 2008, p. 69).

This article reports on the task-based oral production of young EFL learners in Hong Kong where the Ministry of Education officially embraced TBLT in primary schools in the mid 1990s and in secondary schools in 2001 (Carless, 2007). Given the specific cultural teaching context, Hong Kong is in a situation where TBLT is encouraged but not necessarily implemented successfully or used effectively (Carless, 2002, 2003; but see recent initiatives such as the "Primary Literacy Program—Reading and Writing" by the Hong Kong Education Bureau, 2012).

Our study presents an exploratory analysis of the repeated performances of young learners on a spot-the-difference (STD) task. Task repetition, that is, "repetitions of the same or slightly altered tasks—whether whole tasks, or parts of a task" (Bygate & Samuda, 2005, p. 43), has received growing interest in recent TBLT literature (e.g., Bygate, 1996, 2009; Hawkes, 2012; Kim & Tracy-Ventura, 2013; Pinter, 2005, 2007a, 2007b). One of the central issues is how repeating a task affects a learner's language and why. Task repetition is central in Pinter (2005, 2007a, 2007b), who investigates young learners' language by analyzing the shifting interactional patterns and by exploring effects of task repetition on the accuracy and fluency of young learners' performances. Other studies have reported an impact on complexity and fluency (e.g., Bygate, 2009). Yet, as we will explain below, there are theoretical grounds for predicting that repetition could equally well impact on each of

EVELYN SAMPLE & MARIJE MICHEL

the dimensions of **C**omplexity, **A**ccuracy and **F**luency, that is, all three CAF measures (Housen & Kuiken, 2009). Our study deliberately builds on this previous work in an attempt to identify possible trade-off effects between CAF dimensions of L2 performance (Skehan, 2009).

We first review earlier work on task repetition in adult and young learners before we focus on Skehan's (2009) Trade-off Hypothesis. After presenting our study we discuss results in light of the theoretical framework described above and give pedagogical implications and directions for future work.

Task Repetition

Even though earlier work exists (e.g., Plough & Gass, 1993), task repetition as a topic of investigation has been advanced by the work of Bygate (1996, 2009) and Bygate and Samuda (2005). These studies investigated adult learners' repeated performance on different tasks. Bygate (1996) analyzed changes made by a single case regarding the quality of lexico-grammar after a two-day delay. The repeated performance was characterized by higher accuracy, improved use of lexical repertoire, and improved fluency. Bygate's (2009) study among 48 learners of English found that exact repetition led to more complex and more fluent speech. When Bygate and Samuda (2005) examined a subset of 14 participants of the 2001 data, they found the repeated performances to provide more "framing" of events: for example, participants took into account different perspectives and talked more about the intentions of the characters. Stories were more coherent and provided previews, summaries, and explanatory background information. As a whole, these studies made the authors highlight that repetition of exactly the same task made learners shift their attention from message content to more accurate and appropriate formulations (i.e., form) (Bygate, 1996). They concluded that "repeated encounters do not involve the learner in doing the 'same' thing, but rather in working differently on the same material" (Bygate & Samuda, 2005, p. 67). This explanation relates to Levelt's (1989) model of speech production: when repeating a task, learners can draw upon traces of memory specifically related to conceptualization of the task such that meaning generation the second time around requires less attentional resources than during a first encounter. As a result, parallel increases in complexity and fluency may occur.

Other work corroborates the beneficial effects of task repetition on adult L2 performance as it shows steady increases of fluency with mixed findings for complexity and accuracy. For instance, Lynch and Maclean (2000, 2001) report that repeated performances on a poster presentation task promoted complexity and fluency of performance. Birjandi and Ahangari (2008) found complexity and accuracy increased to a minor extent while fluency improved significantly upon task repetition after a one-week delay. Ahmadian and Tavakoli's (2011) results point to gains in complexity and fluency while accuracy revealed no significant changes. The latter authors cite Levelt (1989) and

explicitly relate task repetition to trade-off effects due to limited attentional resources and Skehan's (2009) Trade-off Hypothesis (the shift from meaning to form identified by Bygate, 1996, points in similar directions). We will review the Trade-off Hypothesis below, but first focus on task repetition in young learners.

Task Repetition in Young Learners

Few studies of task repetition have worked with young learners, and the findings seem to differ depending on the construct examined. Stelma and Onat-Stelma (2010) highlight repetition as one important aspect of teaching young learners because of its consolidating function. Van den Branden (1997) investigated 11- and 12-year-olds performing the same task with either a peer or a teacher. The teacher systematically pushed students to provide clear and comprehensible talk. In a posttask repeated performance with a peer, learners who had interacted with the teacher revealed more output and provided more essential information and a wider range of vocabulary while syntactic complexity or accuracy were not differentially affected.

Mackey et al. (2007) studied task repetition in the form of content and procedural task familiarity with 7-to-8-year-olds. Upon repetition, pupils who relied on previously acquired structures were more fluent, whereas learners unfamiliar with task procedures were less accurate and fluent. Children who had performed a similar task before were more sensitive to potential problems and demonstrated more collaborative behaviour than those who had not.

Swain and Lapkin (2008) examined two young immersion learners of French performing a multitask activity that involved repeating a role-play. In the repetition, participants showed a more varied and consolidated use of (new) vocabulary and displayed frequent noticing of new language forms.

Hawkes (2012) studied task repetition as a posttask activity with 13-to-14-year-olds. Students completed the main task then received form-focused instruction before repeating the task. Similar to the results in Van den Branden (1997), Hawkes' students regularly employed the target forms of instruction in the second round, were more accurate, and made more attempts at self-correction. They also revealed greater confidence in using the target language. Following Bygate and Samuda (2005), Hawkes (2012) interprets this as a shift in attention from meaning to form.

Pinter (2005, 2007a, 2007b) looked at task repetition among 10-year-old Hungarian learners of English using a spot-the-difference (STD) task and a direction-giving map task. Repetition made participants more accurate and fluent—complexity was not measured. An in-depth follow-up analysis (Pinter, 2007a, 2007b) showed that, through repeated production, students learned from each other, self-monitored more, and increased in confidence and use of systematic task completion strategies while use of L1 decreased (Pinter, 2005).

EVELYN SAMPLE & MARIJE MICHEL

To recap, previous research seems to agree that task repetition promotes fluency, while offering mixed results for complexity and accuracy. These latter findings, possibly, point toward trade-off effects between these two dimensions (Skehan, 2009).

Task Performance and Trade-off Effects

The Trade-off Hypothesis (Skehan, 2009; earlier known as the Limited Attentional Capacity Model, Skehan, 1998; Skehan & Foster, 2001) proposes that due to capacity limitations, speakers must divide their attentional resources between all the processes a task requires, such as input selection, effective information processing, and response actions. If various task demands exceed the available resources, the ongoing processes and related areas of performance (linguistic complexity, accuracy, and fluency) come into competition with each other. The control function of attention will prioritize certain performative aspects over others (Skehan, 2009): only those aspects receiving enough attention will reach optimal performance while processes under limited attention become erroneous.

VanPatten (1990) argues that learners prioritize meaning over form, as they will try to reach their communicative goal. Consequently, a conflict between form (complexity and accuracy) and fluency will arise (Skehan, 2009): if L2 learners aim at being fluent, less attention will be available for complex and accurate language. A further trade-off is expected between these two latter dimensions because learners may lack resources to pay attention to both complexity and accuracy simultaneously.

Yet, in the case of task repetition, one could expect that trade-off effects decrease as repetitions increase. Hence, the repeated encounters allow L2 performers to shift from meaning-oriented toward more form-oriented production (Bygate & Samuda, 2005). To the best of our knowledge, only two studies have investigated the occurrence of trade-off effects in light of task repetition. Ferrari (2012) performed a longitudinal multiple case study following adolescent learners of Italian performing the same tasks after intervals of up to a year. Effects fluctuated between participants, with some showing a gain in complexity at the cost of accuracy, while others displayed the reverse pattern. Also, Kim and Tracy-Ventura (2013) saw trade-off effects as a possible reason for their results. Their comparison of exact task repetition versus repetition of task procedure by Korean junior high-school learners of English showed no substantial differences between groups. Overall, neither global accuracy nor lexical complexity increased while structural complexity showed some minor growth. The target structure—past tense morphology—did significantly improve in both groups while fluency decreased after repetitions. As the authors indicate, increased attention for the target structure may have resulted in a trade-off with fluency.

The findings of these two latter studies suggest trade-off effects in the context of task repetition, but to date it remains a relatively unexplored domain. The present study tries to fill this gap by investigating Skehan's (2009) Trade-off Hypothesis during task repetition in young learners.

The Present Study

This study is an exploratory examination of six young learners' performances over three repetitions of a spot-the-difference task. We focus on CAF in an attempt to identify trade-off effects in performance (Skehan, 2009). Our research questions are as follows:

RQ1   What effect do three repetitions of a spot-the-difference task have on young learners' complexity, accuracy, and fluency?

RQ2   How do complexity, accuracy, and fluency interrelate with each other over the course of three repetitions of a task?

Participants

Four girls and two boys participated in the study. Their mean age was 9.5 (*SD* = .55), and they had been learning English for at least five years (mean length 6.7 years, *SD* = 1.37). Their oral proficiency ranged from advanced beginner to lower intermediate as judged by one of the authors who was their teacher in an after-school English club for the participants. Participants were selected and paired by this author in consultation with their English teacher, so that they were with a partner they would feel most comfortable with. The club had no effect on participants' regular English lessons.

The Tasks

Four versions of an STD task (see Appendix) were adapted from *Word Bird's Word Book*, a children's picture dictionary (Perrett, 1995). STD tasks are characterized as two-way tasks, encouraging both students to participate equally (Ellis, 2003) and requiring effective collaboration and an understanding of each partner's needs due to an inherent referential challenge (Pinter, 2007a). Task accomplishment depends on successful reference to specific aspects, and exact details are important. As such, STD tasks push participants to be careful to avoid ambiguity or misunderstandings.

Working from one original picture, slightly different pictures were developed for each performance in which six other differences should be identified. Pinter's (2007a) "isolated problem" feature was imperative, that is, none of the differences within each picture were related: if the students missed one difference, they had five other opportunities to be successful in the task. On the one hand, tasks were "new" on each occasion and participants could not simply remember where the differences had been. On the other hand, using the same base picture allowed them to reuse language that was "primed" in

participants' minds from their previous performances (Bygate & Samuda, 2005). Drawing upon procedural and content familiarity (Mackey et al., 2007), the design required students to tackle the same problem and type of language repeatedly within each task. Table 1 illustrates the design.

Table 1
Versions of Tasks Assigned to Student

	Student A	Student B
Performance 1	Picture A	Picture B
Performance 2	Picture B	Picture C
Performance 3	Picture C	Picture D

Tasks were piloted with a different set of pupils and adjusted where necessary. Picture versions and orders were counterbalanced over pairs and performances. The picture versions were distributed to the partners in order to ensure there were exactly six differences between the two pictures in each performance.

Procedure

First, students received a general description and practiced with a different STD task. As they all were willing to participate, a letter was written and translated into Chinese informing the children's parents about the study and requesting written permission for their child's participation.

During the task, each pair of students was asked to identify six differences. Participants recorded their performances themselves. After completion they could gauge their success by comparing each other's pictures. They completed the task outside the classroom during school hours three times within three weeks. Students knew that they would complete the task several times but, in order to prevent them from studying, were not informed exactly when repetitions would be. No similar tasks had been used in their classroom before.

Measures

Task performances were transcribed and, following Housen and Kuiken (2009), coded for measures of structural and lexical complexity (the ability to use advanced language including readiness of risk-taking and use of less controlled language), specific and global measures of accuracy (the avoidance of error in performance including avoidance of challenging structures that may incite error), and fluency (the use of language in real time) as well as task completion using CLAN (MacWhinney, 2000). Table 2 gives an overview of the exact measures.

Table 2

Variables Used to Measure Students' Performances by Means of CAF

Variable	Measurement
Complexity—structural	Clauses/AS-unit[a]
(Foster, Tonkyn, & Wigglesworth, 2000)	Words/AS-unit[a]
Complexity—lexical	Guiraud's Index (Types/√Tokens)
(Skehan, 2009)	D
Accuracy[b]	Error-free clauses/AS-unit[a]
	Agreement errors/AS-unit[a]
	Article errors/AS-unit[a]
	Other errors/AS-unit[a]
Fluency	Filled pauses/minute
(Michel, Kuiken, & Vedder, 2007; Pinter, 2005, 2007a)	Words/minute (per pair)[c]
	Time to complete the task (per pair)
Task completion[d]	Number of differences found

[a] Analysis of Speech units (Foster et al., 2000). [b] We calculated two specific measures, agreement and article errors, because (a) these are frequent sources of errors among the target group of Chinese learners of English (Chang, 2001) and (b) the tasks were designed to elicit correct gender agreement and reference to singular and plural nouns (i.e., task essential structures, e.g., *two girl*, *two girls* is, or masculine/feminine possessive determiners, e.g., *a boy pick up her leg*). All "other" errors (e.g., mistakes in verbal morphology, word order, missing words, and word choice appropriateness) were counted in order to create a global measure (ratio of error-free AS-units) allowing the measurement of overall improvement in accuracy. [c] Unintelligible words and words spoken in the mother tongue were counted as words. [d] Following Skehan (1998), task completion (here: the number of correctly found differences) is more important than any other feature of task performance.

Results

Scores were assessed both overall and within each performance. The following paragraphs present descriptive and inferential statistics as well as correlational analyses to examine trade-off effects.

Task Performance in Terms of Complexity, Accuracy, and Fluency

A general comparison of the means show that complexity values seem to decrease over repetitions although the number of clauses seems to increase; there is no general direction of changes found in accuracy measures; fluency steadily increases over the repetitions, particularly in the second performance. Table 3 displays the means, standard deviations, and range (minimum/maximum) of the three performances for measures of complexity, accuracy, fluency, and the task completion score for all participants.

EVELYN SAMPLE & MARIJE MICHEL

Table 3
Descriptive Statistics for Complexity, Accuracy, Fluency and Task Completion of the Three Performances for All Participants (N = 6)

Measure	Performance 1				Performance 2				Performance 3			
COMPLEXITY	**Mean**	SD	Min.	Max.	**Mean**	SD	Min.	Max.	**Mean**	SD	Min.	Max.
Clauses/AS	**0.84**	0.31	0.44	1.31	**0.93**	0.30	0.45	1.30	**0.93**	0.28	0.62	1.38
Words/AS	**5.43**	1.71	3.30	7.65	**5.19**	1.23	3.34	6.70	**5.01**	1.43	3.08	6.88
Guiraud	**4.32**	0.60	3.59	5.07	**4.27**	0.75	3.05	5.12	**4.11**	0.87	2.94	5.24
D	**19.93**	5.43	13.04	28.27	**18.66**	5.27	9.93	24.93	**17.44**	5.98	8.10	26.02
ACCURACY	**Mean**	SD	Min.	Max.	**Mean**	SD	Min.	Max.	**Mean**	SD	Min.	Max.
Error-free AS	**2.62**	0.65	1.72	3.60	**3.12**	0.96	1.38	4.09	**2.41**	0.51	1.81	3.20
Agreement errors/AS	**0.11**	0.07	0.02	0.20	**0.09**	0.07	0.02	0.16	**0.10**	0.09	0.02	0.25
Article errors/AS	**0.12**	0.09	0.03	0.24	**0.12**	0.05	0.04	0.18	**0.10**	0.04	0.02	0.14
Other errors/AS	**0.40**	0.19	0.00	1.00	**0.47**	0.16	0.00	1.00	**0.52**	0.18	0.00	1.00
FLUENCY	**Mean**	SD	Min.	Max.	**Mean**	SD	Min.	Max.	**Mean**	SD	Min.	Max.
Filled pauses	**29.00**	22.38	5.00	67.00	**14.33**	7.99	6.00	29.00	**15.33**	11.66	5.00	35.00
Words/minute	**43.09**	15.26	24.18	57.31	**51.59**	5.23	46.26	57.85	**53.92**	10.27	45.66	67.03
Completion time	**18.72**	13.06	7.47	35.22	**8.33**	0.92	7.26	9.32	**7.29**	2.33	5.33	10.25
COMPLETION SCORE	**3.83**	0.98	2.00	5.00	**4.50**	0.55	4.00	5.00	**4.83**	0.98	3.00	6.00

Note. AS = Analysis of Speech unit

Complexity

Comparing the raw scores for the complexity measures across the three repetitions, all measures except clauses per AS-unit decrease with repetition. Clauses per AS-unit increase in Performance 2, and stay constant at Performance 3. The raw numbers for Guiraud and D also decrease. The large differences between the minimum and maximum values at each repetition, particularly for lexical complexity (Guiraud and D), suggest individual differences come into play.

Accuracy

Comparing absolute means reveals no general direction for changes in accuracy across the performances. The use of error-free AS-units increases in Performance 2, but decreases in Performance 3 to less than it was in Performance 1. Agreement and article errors per AS-unit both decrease very slightly across the performances. "Other" errors increase with each performance, reflected in the total number of errors (when adding up the three error scores).

Fluency

The means show fluency steadily increasing; improvement is greatest in Performance 2 with fewer pauses, words spoken faster, and a shorter time needed for task completion. In Performance 3 there is slightly more pausing, although faster speech and faster completion time show an overall increase in fluency at repetition three. Again, the minimum/maximum values as well as standard deviations suggest individual variation in fluency.

Completion Score

With each performance, participants were able to detect more differences in the STD tasks than the time before.

Effects of Task Repetition on CAF Measures

Given our non-normally distributed data, we used Wilcoxon Signed Ranks tests to identify any statistically significant differences (alpha-level .05 and effect size r interpreted as small (.1), medium (.3), and large (.5; Cohen, 1988). Considering the restricted value of inferential statistics with small sample sizes, the following figures need to be interpreted with care (e.g., the large effect sizes).

Our data revealed no significant changes over repetition for any complexity measure. A comparison of the first and third performances yielded only a trend: participants tended to show a higher structural complexity for clauses per AS unit in the last performance (z: -1.753, p = .08, mean of ranks: negative = 1, positive = 3.5, r = .51). There was a trend toward lower accuracy from the first to the the third performance measured by the number of other errors (z = -1.782, p = .075, mean of ranks: negative = 2.00, positive = 3.8, r = .51). Simi-

EVELYN SAMPLE & MARIJE MICHEL

larly, the third performance was less accurate than the second—a significant difference when looking at error-free AS units (z = -1.992, p = .046, mean of ranks: negative = 1, positive = 4, r = .58). Participants were significantly more fluent in the second performance than in the first in terms of speed (words per minutes: z = -2.220, p = .026, mean of ranks: negative = 0, positive = 3.5, r = .64). Comparing the first and third performances revealed that their completion score was significantly higher in the third performance (z = -2.121, p = .034, mean of ranks: negative = 0, positive = 3, r = .61). All other comparisons yielded statistically nonsignificant results.

Trade-off Effects

In order to test for trade-off effects between the different dimensions of task performance (Skehan, 2009), Spearman correlational analyses were performed between measures of complexity (lexical: Guiraud and D; structural: clauses/AS unit and words/AS unit), accuracy (total number of errors/AS unit), and fluency (number of filled pauses/AS unit, words/minute, completion time), as well as the completion score (number of differences spotted). All significant correlations showed strong relationships (Spearman's rho > .83). Given the small sample size, these need to be interpreted with caution.

Comparing the first and second performance with the third performance (Table 4), an interesting pattern can be seen: While in the first and second round correlations between structural complexity on the one hand and accuracy and fluency on the other hand were found, this relationship disappears in the third performance. In other words, in the first and second encounter with the task, participants who used complex syntactic structures also made significantly more errors—pointing toward a trade-off between structural complexity and accuracy. In the second performance, lexical complexity also shows significant correlations with the number of filled pauses, that is, participants using an elaborate lexicon were less fluent. Again, this suggests a trade-off between complexity and fluency. Crucially, in the third run no such correlations indicating trade-off effects between CAF-measures arise.

In the first and third performance we see that those who used a more diverse vocabulary (lexical complexity) were most capable of completing the task. In the third performance, they also needed longer for task completion.

Discussion

This study presents an exploratory analysis of six young EFL learners' repeated performance of a spot-the-difference task. It compared the changes in complexity, accuracy, and fluency as well as task completion over three task repetitions and tried to identify trade-off effects among CAF dimensions (Skehan, 2009).

Table 4
Spearman's Correlations Between CAF Measures for Performances 1, 2, and 3

Spearman's Correlations			Str.C wrds/AS	Lex.C Guir.	Lex.C D	A tot. err/AS	F fill. paus.	F wrds/min	F compl. time	TC task compl.
Performance 1										
Structural complexity	clauses/AS	p	.943[b]	.314	.543	.886[a]	.086	-.359	.359	.338
		rho	.005	.544	.266	.019	.872	.485	.485	.512
	words/AS	p		.429	.657	.943[b]	.314	-.239	.478	.338
		rho		.397	.156	.005	.544	.648	.338	.512
Lexical complexity	Guiraud	p			.771	.371	.429	.717	-.359	.845[a]
		rho			.072	.468	.397	.109	.485	.034
	D	p				.486	.543	.478	.120	.845[a]
		rho				.329	.266	.338	.822	.034
Accuracy	total errors/AS	p					.086	-.359	.359	.169
		rho					.872	.485	.485	.749
Fluency	filled pauses	p						.478	.478	.338
		rho						.338	.338	.512
	words/minute	p							-.500	.707
		rho							.312	.116
	completion time	p								-.354
		rho								.492
Performance 2			wrds/AS	Guir.	D	tot. err/AS	fill. paus.	wrds/min	compl. time	task compl.
Structural complexity	clauses/AS	p	.943[b]	.200	.200	1.00[d]	.174	-.239	.239	.488
		rho	.005	.704	.704	n.a.	.742	.648	.648	.326
	words/AS	p		.371	.371	.943[b]	.261	-.239	.239	.488
		rho		.468	.468	.005	.618	.648	.648	.326
Lexical complexity	Guiraud	p			1.00[d]	.200	.899[a]	.717	-.717	-.098
		rho			n.a.	.704	.015	.109	.109	.854
	D	p				.200	.899[a]	.717	-.717	-.098
		rho				.704	.015	.717	.717	.854

EVELYN SAMPLE & MARIJE MICHEL

		fill. paus.	wrds/ min	compl. time	task compl.
Accuracy total errors/AS	p	.174	-.239	.239	.488
	rho	.742	.648	.648	.326
filled pauses	p		.788	-.788	.000
	rho		.063	.063	1.000
Fluency words/minute	p			-1.00[d]	-.408
	rho			n.a.	.422
completion time	p				.408
	rho				.422

Performance 3		wrds/ AS	Guir.	D	tot. err/AS	fill. paus.	wrds/ min	compl. time	task compl.
Structural complexity clauses/AS	p	.754	-.174	-.464	.290	.015	.606	-.667	-.257
	rho	.084	.742	.354	.577	.978	.202	.148	.623
words/AS	p		-.029	-.257	.257	.319	.598	-.478	.169
	rho		.957	.623	.623	.538	.210	.338	.749
Lexical complexity Guiraud	p			.943[b]	.714	.232	-.239	.837[a]	.845[a]
	rho			.005	.111	.658	.648	.038	.034
D	p				.600	.319	-.478	.956[b]	.845[a]
	rho				.208	.538	.338	.003	.034
Accuracy total errors/AS	p					.464	-.359	.359	.676
	rho					.354	.485	.485	.140
filled pauses	p						-.364	.182	.429
	rho						.478	.730	.396
Fluency words/minute	p							-.500	-.354
	rho							.312	.492
completion time	p								.707
	rho								.116

Note. Str. C = Structural complexity; Lex. C = lexical complexity; A = Accuracy; F = Fluency; TC = Task completion; AS = Analysis of Speech Unit; wrds/AS = words per AS; tot. err./AS = total number of errors per AS; fill. paus. = filled pauses; wrds./min = words per minute; compl. time = completion time; task compl. = task completion; [a] = significant at $p < .05$; [b] = significant at $p < .01$, [c] = significant at $p < .001$; [d] The perfect correlation arises due to the fact that Spearman's correlations are based on ranks and the participants of the present study obtained exactly the same rank order for some measures. Identical rho-values for relations with other variables appear.

Task Repetitions by Young Learners

Our first research question asks what effect three repetitions of an STD task have on young learners' complexity, accuracy, and fluency. Our data revealed that task repetition seems to promote task performance, although the three CAF dimensions were not affected to a similar extent and there were fluctuations from the first to the second to the third performance. The fact that we did obtain some significant differences, notwithstanding the small sample size, may allow the following conclusions.

The raw numbers of complexity suggests that students used fewer words but more complicated sentences over the course of the repetitions. Overall there is a (nonsignificant) trend toward more errors and a consistent increase in fluency throughout repetitions, with the largest and most significant improvements in the second run. Students became significantly better at completing the task—that is, with every performance they found more differences.

Our data corroborate earlier work that finds task repetition increases fluency of task performance in adults (Ahmadian & Tavakoli, 2011; Birjandi & Ahangari, 2008; Bygate, 2009; Ferrari, 2012; Lynch & Maclean, 2000, 2001) and younger learners (Mackey et al., 2007; Pinter, 2005, 2007a, 2007b). Mixed findings arise for complexity and accuracy (e.g., Ferrari, 2012; Van den Branden, 1997).

A reason why the students used fewer lexical terms in Performance 2 may be that the language was "primed" from the first performance and students were able to recall lexical items more easily in the following run. Another possibility is that students learned how best to complete the task, using shorter and more direct utterances without lengthy descriptions, also repeating some of what their partner had said. Both are demonstrated in Table 5.

Table 5

Pair C's Performance: Negotiation in Performance 1
vs. Direct Communication in Performance 2

Line	Speaker	
		Performance 1
110	5	have a boy to jump the <tree> [///] sea?
111	6	huh?
112	5	have a boy to jump the sea?
113	6	mm (..) eh where?
114	5	at umm the (..) um on the left.
115	5	is the tree.
116	6	under the tree?
117	5	no. Tree the left. Tree left.
118	6	is eh (..) sit <on the> [//] in the water?
119	5	no.
120	6	jump?
121	5	yes.
122	6	no.
		Performance 2
28	6	eh (.) have a woman is on the water and (.) feel the sunshine.
29	5	feel the sunshine [% said softly] no.
30	6	is near the old woman
31	5	old woman (.) yes.

Note. (.) = 0.5s pause; (..) = 2s pause; <tree> [///] = reformulation; <on the> [//] = retracing with repair; [% said softly] = comment

EVELYN SAMPLE & MARIJE MICHEL

In their first performance, Pair C spends a lot of time negotiating and figuring out what their partner is referring to. The same pair's Performance 2 seems much more fluent and smoother; participants understand what the other is referring to. Participant 5 tends to just softly repeat what his partner has said while scanning his own picture.

Two strategies the students seemed to learn through repetition were to use direct questions and answers, and to provide more precise information to their partner. Both are related to a decreased syntactic complexity—something Ferrari (2012) found in her tasks: as her participants learned the formula for how to best complete the task, they displayed native-like behaviour and their syntactic complexity decreased while accuracy increased. Ferrari concludes that decreasing complexity may actually be regarded as an indicator of development; in other words, "more complex" does not necessarily mean "better" (Pallotti, 2009). Similar strategies in our data can be seen in Table 6.

Table 6
Strategy Use in Pair A

Line	Speaker	
		Performance 2
39	1	(..) mm (..) can you see have a man to (.) eh eh (.) see a riding the bicycle?
40	2	no, I can see.
41	2	I can't see.
42	1	(..) can you see a man is playing football?
43	2	a man?
44	1	yeah, man.
45	2	no, I can't see.
46	1	(..) can you: see a girl is sleeping?
47	1	on the water.
48	2	on the water.
49	1	yes.
50	2	yes, I can see.
51	1	can you see a boy is getting a ball?
52	1	is <on the> [//] at the bottom.
53	2	no, I can see.
54	1	can't see?
55	2	uh I can't see [% "can't" emphasized]

Note. <the girls> [//] = retracing with repair; : = elongated sound; (.) = 0.5s pause; (..) = 2s pause; [% "can't" emphasized] = comment

Participant 1 does most of the questioning, repeating the phrase "Can you see…." Similarly, answers take a uniform format, resulting in a formula through which the pair frames the content of their messages (cf. Bygate, 2013). Participant 1 provides more detailed information to her partner without requiring him to ask (Line 52). This saves time and prevents confusion or misunderstanding.

Limited Attentional Resources and Trade-off Effects

The second research question asks how complexity, accuracy, and fluency interrelate with each other over the course of task repetitions, aiming at identifying

trade-off effects (Skehan, 2009). The central tenet of the Trade-off Hypothesis is that attention devoted to one area may generate a negative impact in another.

As we look at our data, it is noteworthy that Performances 1 and 2 yielded a significant correlation between structural complexity on the one hand and accuracy (and fluency) on the other, which can be interpreted as trade-off effects. The notable finding is that this correlation disappears in Performance 3. In other words, the findings suggest that, in the first two encounters, participants who produced more complex and longer sentences made more errors while those who used simpler syntactic structures made fewer errors. Skehan and Foster (2001) argue that L2 learners under cognitive pressure will prioritize the communicative goal (meaning) of the task over form. Our data may be a result of such prioritization. When students felt the pressure of online processing and devoted attention to keeping a conversation fluent and getting across their meaning, less attention was available for form (either complexity or accuracy). The absence of correlations in Performance 3 points toward an interpretation that participants may not have needed a lot of attention for fluent production anymore—task repetition added to the faster and smoother retrieval of (primed) linguistic information. Consequently, complexity and accuracy were no longer in competition with each other and no trade-off patterns emerged.

Taking a closer look at the correlational analyses gives further support for this interpretation. Skehan (2009) states that learners may not be able to conceptualize and verbalize their message simultaneously while producing fluent speech. "The need to retrieve rarer lexical items seems to have a cost in terms of how error can be avoided and a smooth flow of speech maintained—such retrieval creates processing demands and consumes attentional resources" (Skehan, 2009, p. 518). The first performance presumably shows a focus on message conceptualization: students opted for complex sentences, which correlate with more mistakes. We find a similar relation in the second performance: a higher structural complexity correlates with higher error rates. Lexical complexity, however, yields a trade-off effect with fluency: the more students used the same words (presumably by repeating each other and using sentence frames), the fewer pauses they needed. In the third performance, no trade-offs emerge. Instead, we find a significant improvement in the students' completion score accompanied by a significant correlation between lexical complexity and completion time and score. Most likely, this reflects the fact that as participants spot more differences they will name them and consequently will need more different words (and more time) to do so.

Pinter (2005, 2007a) highlights another aspect: based on the "in-built step-by-step" structure of an STD task, her students became more familiar with the task type and could organize their searches more systematically. Mackey et al. (2007) also noticed this pattern, which Bygate (2013) refers to as "learning the task discourse." Highly structured STD tasks positively affect accuracy and fluency because they allow more attention for ongoing performance

EVELYN SAMPLE & MARIJE MICHEL

(Skehan, 2009) and because the growing familiarity with the task structure facilitates performance over task repetition.

In Performance 1, participants' processing was compromised due to the pressures of the task and working in real time. During the second performance, participants were supported by their first experience: they knew what was needed to complete the task and what to expect from their partner. Not only did participants use language more effectively and provided more essential information (Mackey et al., 2007; Van den Branden, 1997), but they could also draw on the lexical items that were primed by the first performance (Swain & Lapkin, 2008). By the third round, students' strategic behaviour and primed language lessened attentional pressure so that no trade-off effects emerged; the task demands no longer exceeded the resources available.

Evidence of Consolidation

Task repetition may also benefit the consolidation of language (Bygate & Samuda, 2005) as suggested by the excerpts of Pair A's first and second performances (Table 7).

Table 7
Primed Language in Pair A

Line	Speaker	
		Performance 1: Part 1
151	1	can you see a girl is under the tree and the other one the girl to making a … a… flower: … (10)
152	2	yes, I can see.
		Performance 1: Part 2
193	2	(2.5) mm…you can see two girl is (.) take (.) the flower?
194	1	(5.1) no I can't see but I can see the two girl is making (.) a flower +…
195	2	oh, mm (33.5)
196	1	a necklace!
		Performance 2
13	2	(13.9) I can see the man is fishing bottom right [Participant 1 whispering "corner"]
14	1	(19.1) [Participant 1 whispering "corner" twice].
15	1	I can see have two girls is making a flower necklace.
16	1	is under the tree on the right.

Note. (.) = pause of 0.5s; (2.5) = pause for the length of time in parentheses (secs); : = elongated sound; +… = trailing off

In the beginning, Participant 1 could not find the word "necklace" when talking about the two girls. She attempted to use the word twice. She was unable to recall it at first, but spoke it rather triumphantly when she remembered it later. In the second performance, she used the word immediately without hesitation, which suggests she had already internalized the content and words necessary to express her message, relieving some of the pressure felt during the first encounter (Bygate & Samuda, 2005; Swain & Lapkin, 2008).

Evidence of primed language and cognitive reprieve was one of the more noticeable changes in all pairs' performances. Peer assistance (e.g., Participant 1 whispering the word "corner") and strategy development (Bygate, 2013; Pinter, 2007a, 2007b) were other prevalent behaviours in all groups throughout repetitions.

Developing Strategies, Cooperation, Motivation, and Confidence

It appears that all students became more confident in their speech and in approaching the task (Hawkes, 2012; Pinter, 2005, 2007a, 2007b). They adopted strategies by concentrating on certain areas and asking each other for more detail. At first, speakers seemed focused on describing their own picture rather than listening and responding to their partner (again, mirroring Pinter, 2007a, 2007b). In the second and third run, Pair B in particular kept a careful tally of how many differences they had found; see the interaction in Table 8.

Table 8
Strategic Behaviour of Pair B

Line	Speaker	
		Performance 3: Part 1
1	3	um (.) <do you> [///] in the: bottom left corner have two girls are ah (.) saying secret.
2	4	where?
3	3	bottom left corner.
4	4	(..) um. yes.
		Performance 3: Part 2
67	4	(...) on the bottom left corner have a two girl to speaking.
68	3	yes.
69	3	ngo gong jo la [in Cantonese: "I said it already"] how about the girl?
70	3	<the> [///] <have a girl is> [//] (.) um (.) have a girl the trousers is blue.
71	4	yes.
72	4	<the girls> [//] the girl ah the (.) clothes is pink.
73	4	no uh yes is different.
74	3	jong yoh yah go [% in Cantonese: "still one more (difference left)"]
75	3	(...) can ah can you see have a man is take the photo the birds?

Note. <do you> [///] = reformulation; <the girls> [//] = retracing with repair; : = elongated sound; (.) = 0.5s pause; (..) = 2s pause; (...) = 4s pause; A phonetic transcription of words in Cantonese is given with an English translation.

The participants used a strategy they had developed earlier: returning in more detail to areas they had previously discussed in order to spot more differences—a successful strategy because they found something they had missed earlier. Pinter (2007a) reports a similar systematic way to organize searches and find more differences faster.

EVELYN SAMPLE & MARIJE MICHEL

Generally, students' cooperation increased over repetitions, they were able to listen and respond to their partner more effectively, they asked more questions, and they answered each other's questions clearly and in more detail (cf. Bygate, 2013). Students displayed increasingly more negotiation and were able to build joint knowledge and solve ambiguities—a finding that was observed with all the pairs by the third performance. For example, Participant 4 shows a greater understanding of the task and of her partner's needs (Table 8, Line 73): She first says "no," meaning that the girl they are discussing does not have pink clothes. As she understands this may be ambiguous to her partner (she may have been saying "no, this is not a difference"), she immediately clarifies and says "yes is different." Pair A also learns how to best work by explicitly acknowledging previous utterances and asking for clarification (Table 9).

By Performance 3, partners have realized they need to respond to each other. They developed a successful system of asking questions and answers, using the same frames over and again: "can you see…"; "yes, I can see/no, I can't see." These frames presumably account for some of the decrease in lexical complexity found over repetitions.

Table 9
Growing Cooperation by Pair A

Line	Speaker	
		Performance 1
94	2	[% 001 playing with pen] (24.3) I can see a (..) boys playing football in the mable [: middle].
95	1	(6.7) mm, can you see a boy is crying?
96	1	in the middle.
97	2	in the (.) middle [% said very softly] (5.8).
98	1	(34.5) the crying boy is (...) at the top the rainbow (..) on the left.
99	2	(18.1) mm (57.0) [% 001 sigh] (23.8) you can see the cow is <o on> [: on] the left?
		Performance 3
9	1	can you see a man to sleeping on the hill?
10	2	mm?
11	1	he on the right.
12	2	no, I can't see.
13	1	can you see a boy (.) is playing football?
14	2	sorry, <I can> [//] I can't see.
15	2	<you> [//] can you see the cow?
16	2	on your left.
17	2	yes, I can see.

Note. [: middle] = replacement; [% sigh] = comment; (24.3) = 24.3s pause; (.) = 0.5s pause; (..) = 2s pause; <o on> = replacement; [//] = retracing with repair

It is interesting to compare this pair's excerpts from Performance 1 with Performances 2 and 3 (cf. Tables 6 and 9). At first, there were frequent long pauses (in Line 99 taking almost a minute)—supporting the idea of a trade-off in favor of a lexical search at the cost of fluency. Participant 2 was soft-spoken and stumbled in his speech while his partner was getting frustrated during the time she had to wait (whispering answers, sighing). In their third performance (see Table 9) the pair was much more fluent. Participant 2 did not seem hesitant anymore. Both students were using words and structures they had encountered before that, therefore, presumably required fewer cognitive resources. The interaction appears to flow well and be efficient because participants understood and responded to each other's needs. Similarly, growing task familiarity seemed to add to effective task completion (Mackey et al., 2007).

Finally, in line with Hawkes (2012) and Pinter (2007a), we found that task repetition positively influences affective variables such as motivation and confidence. For example, each difference identified that contributed to success and task completion raised enthusiasm as demonstrated by Participant 3 (cf. Table 8, Lines 74–75), who mentions they only have one more difference left and immediately launch into another specific and direct question in order to find the last one. Participant 2 demonstrated increased confidence: in the third round, it was him turning on the recorder after asking his partner "Are you ready?" Up to then, his partner had been doing this. Furthermore, this time Participant 2 provided more details, such as exactly where a golfer was, right from the beginning, which he had been reluctant to do in the first performance.

Pedagogical Implications

Even though this study presents data from a small out-of-class sample, some pedagogical implications may be drawn. Our data suggest that students benefitted from task repetition, although in different ways. As such, repeating a task could be manipulated in order to achieve specific pedagogic goals.

For example, if fluency is the target, only one repetition may be enough. In our study the most significant gains in fluency were found between Performances 1 and 2. In contrast, accuracy and complexity may benefit more through several repetitions given the fact that trade-off effects between these dimensions disappeared over time. It may be fruitful for teachers to consider trade-off effects when using or assessing by means of tasks. Hence, learners seemed to prioritize a specific CAF dimension to deal with task demands. Accordingly, task repetition may help to achieve a more balanced performance over time.

From a classroom perspective it is noteworthy that students did not mind repeating the task. Rather, their enjoyment grew over repetitions. Allowing learners to stretch to their upper potential is an important pedagogic goal as it adds to motivation and confidence. Using slightly altered versions of tasks (rather than exact repetition, cf. Kim & Tracy-Ventura, 2013) may be

a way for them to draw on prior knowledge of content and task procedure, enabling more active participation and generating more opportunities for development (Mackey et al., 2007) without losing motivation. Overall, task repetition—especially of STD tasks—presents opportunities for students, as it supports both interaction with one another and target language use without teacher interference.

Limitations and Directions for Future Research

We are aware that we draw on a very small sample of six students who participated in only one task in an out-of-class context. Students' individual behaviour may well have influenced the overall picture, and our study cannot claim a direct cause-effect relationship between our findings and task repetition. Moreover, the task performances were collected over three weeks, during which time it would be unnatural for students not to share or discuss their experiences with others—which probably affected their next performance.

Future work with young learners would benefit from a larger sample and could elaborate on the existing studies by using more and/or other task types. A longer sequence of repetitions (more than three performances) could be administered to investigate the fluctuation of repetition effects and determine how long the impact of repeated performance lasts. Additionally, the longer-term effects of repetition on performance and development could be measured with a longitudinal study.

In sum, the present work can only be considered exploratory. Nonetheless, based on our data we have formulated several hypotheses (e.g., pointing towards trade-off effects; Skehan, 2009) that are worth testing in the future.

Summary and Conclusion

The present study explored young EFL learners' complexity, accuracy, and fluency over repeated performance of a spot-the-difference task targeting the identification of trade-off effects. Mirroring earlier work on task repetition, we found a consistent and significant increase of fluency through task repetition and a mixed picture concerning complexity and accuracy. We have shown that in the initial performances, the benefits on one dimension of task performance (e.g., structural complexity or fluency) seemed to come at the expense of a trade-off with another dimension (e.g., accuracy or lexical complexity). By the third performance, however, these trade-off effects disappeared, suggesting that as the students became familiar with the task they were able to focus their attention on all three CAF areas simultaneously (Skehan, 2009).

As many other positive aspects of task repetition were also found, including increased cooperation, motivation, confidence, and strategy development, we conclude that task repetition is a valuable pedagogic tool for young learners given that they benefitted in ways that were specific to them and their needs.

Note

1   Following Ellis (2014), our participant group would be more specifically classified as primary school pupils.

Acknowledgements

The Authors

Evelyn Sample received her Master's in TESOL from Lancaster University. She also holds an MA in Sociology and Social Anthropology from the University of Edinburgh. She is currently teaching English as a foreign language in Hong Kong.

Marije Michel holds a PhD in Applied Linguistics from the University of Amsterdam. Since 2012 she has been a lecturer for Language Learning and Teaching at Lancaster University, UK. She is interested in cognitive and interactive aspects of second language acquisition, in particular task complexity as well as computer-mediated communication.

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, *15*(1), 35–59.

Birjandi, P., & Ahangari, S. (2008). Effects of task repetition on the fluency, complexity and accuracy of Iranian EFL learners' oral discourse. *Asian EFL Journal, 10*(3), 28–52.

Bourke, J. M. (2006). Designing a topic-based syllabus for young learners. *ELT Journal, 60*(3), 279–286.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). Oxford, UK: MacMillan Heinemann.

Bygate, M. (2009). Effects of task repetition on the structure and control of oral language. In K. Van den Branden, M. Bygate, & J. Norris (Eds.), *Task-based language teaching: A reader* (pp. 249–274). Philadelphia, PA: John Benjamins.

Bygate, M. (2013, October). *On fetters and goals, and the development of an empirical TBLT in terms of language, learning and teaching*. Plenary speech presented at the 5th Biennial International Conference on Task-based Language Teaching, Banff, AB, Canada.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37–74). Philadelphia, PA: John Benjamins.

Carless, D. (2002). Implementing task-based learning with young learners. *ELT Journal, 56*(4), 389–396.

Carless, D. R. (2003). Factors in the implementation of task-based teaching in primary schools. *System, 31*(4), 485–500.

Carless, D. (2007). The suitability of task-based approaches for secondary schools: Perspectives from Hong Kong. *System, 35*(4), 595–608.

Chang, J. (2001). Chinese speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd ed., pp. 310–324). Cambridge, UK: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Ellis, G. (2014). "Young learners": Clarifying our terms. *ELT Journal, 68*(1), 75–78.

Ellis, R. (2003). *Task-based language learning and teaching.* Oxford, UK: Oxford University Press.

Ferrari, S. (2012). A longitudinal study of complexity, accuracy and fluency variation in second language development. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 277–297). Philadelphia, PA: John Benjamins.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354–375.

Hawkes, M. L. (2012). Using task repetition to direct learner attention and focus on form. *ELT Journal, 66*(3), 327–336.

Hong Kong Education Bureau. (2012). *NET scheme in primary schools: Primary literacy programmes.* Retrieved from http://www.edb.gov.hk/en/curriculum-development/resource-support/net/plpr.html

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461–473.

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System, 41*(3), 829–840.

Levelt, W. J. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research, 4(3)*, 221–250.

Lynch, T., & Maclean, J. (2001). A case of exercising: Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 141–162). Harlow, UK: Longman.

Mackey, A., Kanganas, A. P., & Oliver, R. (2007). Task familiarity and interactional feedback in child ESL classrooms. *TESOL Quarterly, 41*(2), 285–312.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Michel, M., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching, 45*(3), 241–259.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590–601.

Perrett, J. (1995). *Word bird's word book: Vocabulary building for young learners.* Hertfordshire, UK: Prentice Hall International.

Pinter, A. (2005). Task repetition with 10-year-old children. In C. Edwards & J. Willis (Eds.), *Teachers exploring tasks in English language teaching* (pp. 113–126). Basingstoke, UK: Palgrave Macmillan.

Pinter, A. (2007a). Some benefits of peer-peer interaction: 10-year-old children practising with a communication task. *Language Teaching Research, 11*(2), 189–207.

Pinter, A. (2007b). What children say: Benefits of task repetition. In K. Van den Branden, K. Van Gorp, & M. Verhelst (Eds.), *Tasks in action: Task-based language education from a classroom-based perspective* (pp. 131–158). Newcastle, UK: Cambridge Scholars Publishing.

Plough, I., & Gass, S. M. (1993). Interlocutor and task familiarity: Effects on interactional structure. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning: Integrating theory & practice* (pp. 35–52). Clevedon, UK: Multilingual Matters.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning.* Hampshire, UK: Palgrave Macmillan.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford, UK: Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge, UK: Cambridge University Press.

Stelma, J., & Onat-Stelma, Z. (2010). Foreign language teachers organising learning during their first year of teaching young learners. *Language Learning Journal, 38*(2), 193–207.

Swain, M., & Lapkin, S. (2008). Lexical learning through a multitask activity: The role of repetition. In T. W. Fortune & D. J. Tedick (Eds.), *Pathways to multilingualism: Evolving perspectives on immersion education* (pp. 119–132). Clevedon, UK: Multilingual Matters.

Van den Branden, K. (1997). Effects of negotiation on language learners' output. *Language Learning, 47*(4), 589–636.

VanPatten, B. (1990). Attention to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*(3), 287–301.

Appendix
Examples of a Spot-the-Difference-Task
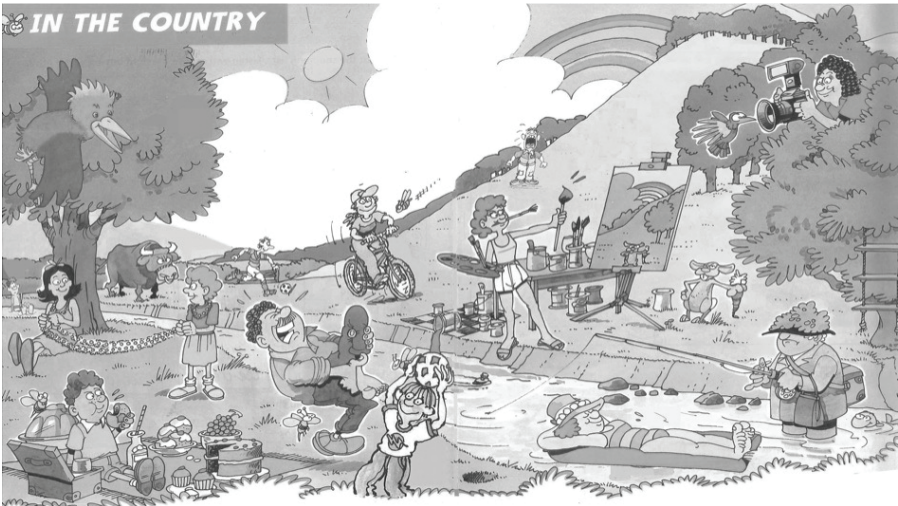
Pictures adapted with permission to publish from Perrett (1995).
The full set of pictures is available on www.iris-database.org.