Are two heads better than one? Pair work and grammatical accuracy

N. Storch*

Centre for Communication Skills and ESL, The University of Melbourne, 138 Cardigan Street, Carlton, Victoria 3053, Australia

Received 1 October 1998; received in revised form 15 December 1998; accepted 15 January 1999

Abstract

The use of pair work has been promoted in both first (L1) and second (L2) language classrooms. In the L2 classroom, a number of studies have shown that learners working in pairs have more opportunities to communicate in the target language than in teacher-fronted classrooms. However, this research has also shown that the tasks generally used in such studies (e.g. jigsaw) do not engage students in negotiations over grammar. In the language class where the development of both fluency and accuracy are important goals, what is needed is research on grammar-focused communication tasks investigating the effects of student negotiations over grammatical choices on the accuracy of production. The small-scale study reported here required tertiary ESL learners of intermediate to advanced L2 proficiency to complete three different types of grammar-focused exercises commonly used in the language classroom: a cloze exercise, a text reconstruction and a short composition. Each exercise type had two isomorphic versions, one to be completed individually and the other to be completed in pairs. A comparison of exercises completed individually with those completed in pairs suggested that collaboration had a positive effect on overall grammatical accuracy, but tended to vary with specific grammatical items. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Pair work; Grammar tasks; Metatalk

1. Introduction

Research findings in both first (L1) and second (L2) language learning have long been supportive of the use of small groups and pair work in the language classroom.
Studies in L1 pedagogy have shown that learners working in groups are exposed to a variety of viewpoints, co-construct new ways of understanding and develop critical thinking skills (Adams and Hamm, 1996; Barnes and Todd, 1977; Slavin, 1991). In the field of L2 pedagogy, Long and Porter’s (1985) review of studies on group work presents a number of pedagogical and psycholinguistic arguments supportive of group work. Group work has been shown to provide L2 learners with opportunities to use the target language for a range of functions. Furthermore, compared to teacher-fronted classes, learners in groups have been shown (eg. Pica and Doughty, 1985; Doughty and Pica, 1986) to engage in more conversational modifications, or what Long (1983) refers to as “negotiations of meaning”. These negotiations are said to make input more comprehensible and hence facilitate L2 acquisition.

However, as Pica (1994, 1997) admits, what most of the L2 studies to date have shown is that learners negotiate most readily on lexical items and larger syntactic units, whereas negotiation over grammatical morphology is rare. Pica (1997) also notes that most of the cited studies have utilised tasks which aim to provide learners with genuine oral communication practice. It thus seems that such tasks do not succeed in drawing learners’ attention to grammatical forms and structures in need of further development (eg. verb tense choice). Pica calls for more studies which utilise tasks designed to tap into grammatical modifications.

Recently, researchers such as Kowal and Swain (1994) and Swain and Lapkin (1998) have proposed the use of tasks which require collaborative written output as a means of drawing learners’ attention to grammatical accuracy within a communicative context. These researchers argue that such collaborative output tasks may be of benefit, particularly in developing learners’ grammatical competence, because the joint act of production generates metatalk and reflection about language choices. Such verbalisation and reflection may serve to raise learners’ attention to ‘gaps’ in their knowledge or provide them with positive feedback and thus promote language learning. However, Swain (1998) also found that learners working in pairs not only learn from metatalk leading to correct decisions but that they also internalise incorrect decisions.

Thus, the question of whether metatalk leads to more accurate grammatical decisions is an important one, but one that has not been fully investigated. Results from some studies suggest that metatalk may have a positive effect only on the accuracy of a few individual grammatical structures. For example, Goss et al. (1994), who compared grammatical judgements tasks completed individually and in pairs, found modest differences in favour of dyads on some grammatical features such as referential pronouns. Another study of relevance here is by Stratman and Hamp-Lyons (1994) who investigated the effects of individual metatalk—think alouds—on grammatical accuracy. The researchers compared students’ products (text revision) under two conditions: with and without think aloud. Their findings suggested that the verbalisation in the think aloud may have led to greater accuracy in the use of pronouns but that it had little impact on improving other morphosyntactic errors such as verb tense.

The study reported here is part of a larger study on collaborative pair work. The study set out to investigate the nature and effect of metatalk on the accuracy of
grammatical decisions. Findings concerning the nature of the metatalk are reported elsewhere (Storch, 1998). This paper reports on findings concerning the effect of pair work and the associated metatalk on grammatical decisions, and addresses the following two research questions: Do students working in pairs and discussing their grammatical choices produce more accurate written texts than students working on similar exercises individually? And if so, do all grammatical items benefit from such collaboration?  

2. The study  

2.1. The participants  

The participants in this study were graduates from an English for Academic Purposes (EAP) course offered by the ESL program at a large Australian university. Eleven students (two male, nine female) volunteered to participate in the study, but only eight completed all required tasks. The participants came from a range of language backgrounds (Indonesian, Chinese, Japanese, French, Russian, Spanish, Persian). The mean length of residence in Australia was 3.5 years. The students’ ESL proficiency, based on in-house ESL placement tests, ranged from intermediate to advanced.  

2.2. The language exercises  

Three different language exercises were used in this study. These exercises were chosen because they represent the kind of exercises students may encounter in language or grammar classes. The three exercises were; a cloze exercise, representing perhaps the more traditional grammar-based exercises; and a text reconstruction and composition, representing tasks which involve students in producing written output, and which are the kinds of tasks advocated by researchers such as Kowal and Swain (1994) and Swain and Lapkin (1998). Each exercise had two versions, one version to be completed by each student individually, the other version to be completed in pairs (or small groups). As far as possible, the two versions were isomorphic: they were on the same theme, of the same genre, of approximately the same length and required attention to approximately the same number of similar grammatical items (Appendix).  

The grammatical items that were focussed on included articles (definite and indefinite), verb tense/aspect choice and formation (in particular, simple present, simple past and present perfect), derivational morphology (adjectival, nominal or adverbial suffixes) and nominal morphology (possessive or plural marking). These grammatical items were chosen because they were found to be particularly inaccurate in the

---

1 The term ‘collaboration’ is used throughout this paper to refer to students working in pairs or small groups, completing a set exercise jointly. However, the author acknowledges that the term ‘collaboration’ means much more than mere joint production and indeed the larger study aims at explicating this term.
students’ essays on the placement tests. Other grammatical items, such as prepositions and linking devices (inter- and intrasentential links) were also included but to a lesser extent.

2.3. Procedure

The study required students to attend two sessions. In the first session students completed a cloze exercise and a composition individually. Then, following a brief demonstration of a text reconstruction task, students completed a text reconstruction task in pairs. In the second session (2 days following the first session), a text reconstruction task was completed individually and a cloze exercise and composition were completed in pairs.

In the collaborative mode, students worked in self-selected dyads (and one triad because of uneven student numbers). In order to encourage joint production, each dyad (or triad) was given only one copy of the task. All pair talk was audio-taped and the researcher recorded the time it took the participants to complete the set tasks. The researcher noted that when students worked in pairs, the amount of time spent on completing the exercises compared to individual efforts almost doubled.

Did this increase in response time spent on completing the set tasks and the fact that students worked in pairs have an effect on grammatical accuracy? To examine the possible effect on grammatical accuracy of pair work and the verbalisation it entails, a comparison of the texts completed individually and those completed collaboratively was undertaken.

3. Findings

For the purposes of this paper, the main source of data was the completed language exercises. The following sections outline how the text of each language exercise was analysed for grammatical accuracy of the targeted items and for overall accuracy, how the two versions were compared and present the findings from the analyses.

3.1. Cloze exercise

When marking the cloze exercises, an acceptable answer scoring system was utilised. Each version had 16 gaps; however, the number of tokens for each of the targeted grammatical items differed slightly. Average accuracy scores (in percentages) for the targeted grammatical items only (excluding prepositions, for which there was only a single item in each version) and for overall accuracy for version 1 (completed by nine individuals) and for version 2 (completed by four dyads) are shown in Table 1.

On the basis of the descriptive statistics shown in Table 1, the cloze exercise completed by pair work was more accurate (an increase in the average total score from
58 to 77%). This improved accuracy occurred in decisions concerning verb tense/aspect choices (up from 58 to 78% accurate) and particularly in decisions over derivational morphology (up from 35 to 84% accurate responses). In contrast, in the case of articles, collaboration seemed to produce a fall in accuracy: from a high average score of 94% when completed individually to an average of only 63% correct choices when completed collaboratively. However, given the small number of tokens for article use particularly in version 1 of the cloze exercise, these results need to be interpreted with caution.

3.2. Text reconstruction

Comparing accuracy scores for the two versions of the text reconstruction task was more difficult than for the cloze exercise. As a task which entails some production, there were a number of ways students could reconstruct many of the sentences. For example, when presented with the proposition “This article report on study...” (Appendix), possible response patterns included: inserting a correct article (on a study); detecting the need for an article but supplying an incorrect article (on the study) or failing to note the need for an article altogether (on study). Also, unlike the multiple-choice and cloze exercises, where students were constrained by the choices offered or the gaps, in the text reconstruction task participants could add words. Thus, in the above example, some students reconstructed the proposition by inserting a new verb and an additional article, as in “This article gives a report on a study...”. This also led to variations in the number of tokens per grammatical item for each version.

Thus, the products of the text reconstruction task were analysed firstly for all possible response patterns including the ability to detect the need for an amendment and the ability to amend correctly (see also Stratman and Hamp-Lyons, 1994, on analysing students’ editing exercises). Furthermore, all additions and deletions, whether acceptable or unnecessary, were noted. It was only then that comparison for individual grammatical items was undertaken.

Table 2 summarises the findings for the text reconstruction task completed under the two conditions. The table shows the various possible responses as a percentage of total decisions made for the task. The final row shows the number of correct decisions for each task.
decisions (detected and correctly amended as well as any acceptable additions) as a percentage of total decisions taken in completing the task.

Table 2 shows that, overall, having students work on the text reconstruction in pairs had a positive effect on the task performance. Collaboration seemed to have a positive effect on the participants’ ability to detect the need for an amendment and to provide the correct amendments. A greater proportion of items were detected and correctly amended (72% compared to 63%) and fewer were left undetected (10% compared to 17%). Indeed, a greater proportion of all decisions made by the pairs were correct (74%) compared to decisions made by the students individually (65%).

Reconstructed texts were also analysed for the targeted grammatical items amended or added. For each version of the task, the number of correct decisions for each grammatical item was counted, and expressed as a percentage of total decisions made for that grammatical item. Table 3 shows the average correct score for each of the targeted grammatical items for version 1 (completed by eight individuals) and for version 2 (completed by three dyads and one triad). Given the unequal number of tokens for each grammatical item, the range of tokens for each grammatical item is included in Table 3.

Again, the findings show that having students work in pairs seemed to have a positive effect on grammatical decisions. A greater percentage of decisions concerning all the targeted structures were correct when the text was reconstructed in pairs.

Table 2
A comparison of text reconstruction exercises completed individually and collaboratively

<table>
<thead>
<tr>
<th>Possible responses expressed as % of total changes made</th>
<th>Individually (n = 8)</th>
<th>Collaboratively (n = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items detected and correctly amended</td>
<td>63</td>
<td>72</td>
</tr>
<tr>
<td>Items detected but incorrectly amended</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Undetected</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>Unacceptable additions/deletions</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Acceptable additions/deletions</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Correct decisions</td>
<td>65</td>
<td>74</td>
</tr>
</tbody>
</table>

Table 3
A comparison of reconstructed texts completed individually and collaboratively per grammatical item

<table>
<thead>
<tr>
<th>Average score (%)</th>
<th>Articles a</th>
<th>Verb tense/aspect b</th>
<th>Derivational morphology c</th>
<th>Nominal morphology d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version 1 (completed individually)</td>
<td>67</td>
<td>65</td>
<td>55</td>
<td>63</td>
</tr>
<tr>
<td>Version 2 (completed collaboratively)</td>
<td>77</td>
<td>77</td>
<td>67</td>
<td>86</td>
</tr>
</tbody>
</table>

a Version 1, n = 8–12; version 2, n = 8–9.
b Version 1, n = 8–10; version 2, n = 8–9.
c Version 1, n = 2–4; version 2, n = 2–3.
d Version 1, n = 3; version 2, n = 3–4.
However, unlike the cloze exercise, in the text reconstruction, modest gains in accuracy occurred in all the targeted grammatical items: articles, verb tense/aspect, derivational morphology and with the greatest gain in decisions concerning nominal morphology (an increase in average accuracy from 63% when completed individually to 86% when completed collaboratively).

### 3.3. Composition

Comparing the composition text written under the two conditions proved to be the most complex analysis. The compositions produced varied in length and syntactic complexity and hence accuracy scores needed to take these variations into consideration. Thus, length of compositions (in terms of words), number of sentences and clauses and complexity measures such as the ratio of clauses to sentences\(^2\) were calculated prior to analysing compositions for grammatical accuracy. Then, in order to measure differences in accuracy, the total number of errors, the type of errors made and the ratio of error-free clauses to the total number of clauses produced were calculated. Since the focus of this study was on grammatical accuracy, errors in spelling, vocabulary choice and punctuation were ignored in the analysis. Table 4 presents the comparison between the compositions under the two conditions: individually and collaboratively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Individually ((n = 8))</th>
<th>Collaboratively ((n = 4))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Length</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average No. of words</td>
<td>177 (range: 126–264)</td>
<td>162 (range: 138–187)</td>
</tr>
<tr>
<td>Average No. of sentences</td>
<td>10.75 (range: 5–13)</td>
<td>11.25 (range: 5–11)</td>
</tr>
<tr>
<td>Average No. of clauses</td>
<td>19</td>
<td>14.5</td>
</tr>
<tr>
<td>Average length of sentences (in words)</td>
<td>16.45</td>
<td>14.4</td>
</tr>
<tr>
<td><strong>Complexity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average clause/sentences</td>
<td>1.78</td>
<td>1.29</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average No. of errors</td>
<td>13.6</td>
<td>7.75</td>
</tr>
<tr>
<td>Error-free clauses/total No. of clauses (%)</td>
<td>47</td>
<td>61</td>
</tr>
</tbody>
</table>

\(^2\) A common measure of linguistic complexity (see Gaies, 1980, for discussion) used in the 1970s and 80s was the ratio of clauses to the T-unit. However, a number of researchers have criticised the use of T-units as an index on both definitional and procedural grounds and have questioned its validity (Gaies, 1980). Indeed, with more advanced learners, Bardovi-Harlig and Boffman (1989) claim that the sentence rather than T-units is a more appropriate unit of measure of syntactic complexity because sentences capture facts about conjunctions and parentheticals which a T-unit misses. Thus, given that the participants in this preliminary study were of intermediate to advanced L2 proficiency, the number of sentences and clauses—both dependent and independent—were counted.
Table 4 shows that the pairs produced shorter and less syntactically complex compositions. Compositions produced individually ranged from 126 to 264 words whereas most compositions produced collaboratively were shorter, ranging from 138 to 187 words and averaging 162 words. Although compositions produced under the two conditions had, on average, a similar number of sentences, the number of clauses per sentence and words per sentence differed. Sentences in essays produced individually tended to be longer (over 16 words in length) and had more clauses (19 compared to 14.5 in collaboratively produced composition). Furthermore, on average, compositions produced in pairs scored lower complexity scores in terms of ratio of clauses to sentences (an average of 1.78 clauses per sentence when produced individually to only 1.29 when produced collaboratively). However, collaboration had a positive effect on accuracy. Compositions produced collaboratively in pairs demonstrated a lower average number of errors (7.75 compared to 13.6 errors) and a greater proportion of error-free clauses (61% of error-free clauses in compositions produced collaboratively compared to 47% when produced individually). There were no clear differences in the type of grammatical errors made in compositions under the two conditions.

4. Discussion

The analysis showed that, overall, when students completed tasks in pairs their joint effort was more accurate. Unlike the results of the study by Stratman and Hamp-Lyons (1994) mentioned earlier, which found that the process of verbalisation associated with think aloud did not aid learners in detecting word-level errors, this did not seem to be the case in this study. In all the language exercises in this study, collaboration and the metatalk it generated led to an improvement in the grammatical accuracy of the texts produced.

This improved accuracy may be due to the longer time students spent on the exercises when completing them in pairs. As noted earlier, the time taken to complete the exercises in pairs almost doubled when compared to the time taken to complete them individually. Indeed, transcripts of pair talk and the observation notes taken indicate that when working collaboratively, students seemed more motivated to focus on grammatical accuracy. That is, the additional time taken to complete all the tasks collaboratively was due to the fact that students tended to revise their product a number of times. Whereas when working individually, students tended to complete their work quickly and very few revised their work before submitting it to the researcher.

In the composition task, it could be argued that accuracy gains were at the expense of linguistic complexity. In particular, in the case of the composition, the collaboratively produced texts were shorter and less linguistically complex. A study by Hulstijn and Hulstijn (1984), which involved participants in an oral retelling task, also found that focusing learner attention on form resulted in higher accuracy but reduced the information content of the product.

Greater grammatical accuracy may have also been due to what Stratman and Hamp-Lyons (1994) as well as Russo et al. (1989) refer to as the acoustic or auditory
feedback available during verbalisation. In this study, pair work provided opportunities for two types of feedback: individual acoustic feedback as learners verbalised their own decisions, as well as for peer feedback.

However, in terms of the second research question, i.e. whether collaboration and the associated metatalk led to more correct decisions for the targeted grammatical items, the findings seem inconclusive. In the case of articles in particular, it is unclear whether metatalk necessarily led to greater correct decisions. Article choices in the reconstructed texts were more accurate when the task was completed in pairs but not so in the case of the cloze task. Yet for decisions concerning derivational morphology collaboration increased accuracy in all exercises. One possible explanation for these findings is that perhaps less complex grammatical items, such as derivational morphology and nominal morphology, where decisions can be based on clear rules, are more amenable to conscious verbalised reflections. On the other hand, on grammatical items such as articles, which are renowned for the complexity of rules governing their use, conscious reflection and deliberation may lead to incorrect decisions.

There also seemed to be some relationship between the type of language exercise and the grammatical accuracy of specific grammatical items. In the case of articles discussed above, collaboration reduced the accuracy of article use on the cloze exercise—a language exercise with a more overt focus on grammatical decisions and which forced students to focus attention on the targeted structures (given the constraints of gaps). On tasks which required some production (text reconstruction and composition) and where students had more choice over which grammatical decisions to discuss (or as in the case of composition not to discuss; Storch, 1998), article choices were more accurate. It may be that coercing learner metatalk on certain complex structures, such as articles, may indeed lead to incorrect choices. Given the small-scale nature of this study, these findings are suggestive and further studies are needed which examine the interaction of task type and grammatical structures.

5. Conclusion

The findings from this small-scale study suggested that having students complete language exercises in pairs may have mixed effects on grammatical accuracy: a positive effect on the overall grammatical accuracy of a text but a varying effect on certain grammatical items. These findings would therefore suggest that perhaps not all grammatical items and structures benefit from the same kind of classroom treatment.

However, it is important to emphasise that these findings are suggestive, given the small scale of this study and number of tokens per grammatical items and that therefore further research is necessary in classroom settings to investigate the effect of pair work on grammatical accuracy. Given the widespread use of pair work in language classrooms, and the interest in developing language exercises and tasks which can develop L2 learners’ grammatical accuracy, such research seems pertinent.
Furthermore, it should be noted that this study investigated the effect of collaboration on production. It did not address the larger question, i.e. the effect of collaboration and metatalk on the development of learners’ grammatical accuracy in the longer term.

Appendix: Tasks used in the study

A.1. Cloze

A.1.1. Version 1
Instructions: Please complete the following task on your own. Complete the following text by filling in the gaps with an appropriate word (e.g. article, verb, prepositions, etc.).

The immigration policy of _______ country represents its _______ to a number of important policy questions. Throughout Australia’s history, the two main issues in the immigration debate _______ the size and composition _______ the immigration program.

Traditionally, Australia’s migrants _______ from the United Kingdom and Ireland. In _______ 1980s, the number of migrants from these countries _______. In 1981, they _______ 31.6 per cent of settler arrivals, but in 1991 this contribution _______ to 16.2 per cent. At the same time, the number of settlers from other _______ countries, such as Italy, Greece and Yugoslavia also _______. However, despite this _______ from the more _______ sources, the overall number of settlers in Australia _______ substantially during the last decade due to the _______ in settlers who _______ from Asia.

A.1.2. Version 2
Instructions: Please complete the following task with a partner. Complete the following text by filling in the gaps with an appropriate word (e.g. article, verb, prepositions, etc.).

The Australian government determines its immigration policy on _______ annual basis. The program _______ people who come _______ Australia on an immigrant visa and those who _______ for a _______ stay but successfully apply to remain permanently.

The immigration program distinguishes between _______ categories of immigration including Family, Skilled, Humanitarian and Special Eligibility. During _______ late 1980s, Family migration _______ a major component of the immigration program. However, _______ actual number of migrants who _______ under this category in 1990 _______ by approximately 10,000. This _______ was due to changes in immigration processing arrangements.
In 1988–9, the number of settlers who ________ under Skilled migration ________ down slightly. However, it ________ substantially in 1989-90 as a result of a new point system ________ government implemented in 1989.

A.2. Text reconstruction

A.2.1. Version 1
Instructions: Please work on the following task on your own. Reconstruct the following text inserting all the necessary grammatical words (e.g. articles, prepositions, linking words, etc.), and changing verbs where necessary to produce a meaningful and grammatically correct paragraph.

This article report on study carry out Professor Graeme Hugo Adelaide University. Study investigate emigration trends and economy implications of emigration. Study find many emigrant people with above average profession qualifications. Last year, trend reach new post-war peak. Professor Hugo reject view that Australia immigration program fail. He say that skill migrants still general outnumber emigrants.

A.2.3. Version 2
Instructions: Please complete the following task with a partner. Reconstruct the following text inserting all the necessary grammatical words (e.g. articles, prepositions, linking words, etc.), and changing verbs where necessary to produce a meaningful and grammatically correct paragraph.

This article report on study carry out Professor Graeme Hugo Adelaide University. Study investigate emigration trends and economy implications of emigration. Study find most people leave Australia since 1947 former migrants. Many migrant leave when retire. Report attribute finding to ability of Australia to receive pensions overseas. For example, until quite recent, over half all Australia pensions send Greece and Italy.

A.3. Composition^3

A.3.1. Version 1
Instructions: Write a short text (2–3 paragraphs) in which you describe what you do in AESL, language class 1, to a new student who is not familiar with the course.

---

^3 It should be noted that the composition topics required students to describe very different language classes. In language class 1 (the topic for version 1 of the task), the class focus, and hence of the exercises, is on developing learners’ academic listening and speaking skills. In language class 2 (the topic for version 2 of the task), the focus is on developing learners’ academic reading and writing skills.
A.3.2. Version 2

Instructions: In pairs, write a short text (2–3 paragraphs) in which you describe what you do in AESL, language class 2, to a new student who is not familiar with the course.

References


