Speaking, writing, and memory span performance: Replicating the Bourdin and Fayol results on cognitive load in German children and adults

Joachim Grabowski

on leave from University of Education at Heidelberg (Germany)
presently at Catholic University of Eichstätt (Germany)

grabowski@ph-heidelberg.de

Keywords: working memory, cognitive load, automation, writing development, spelling

Abstract

In an influential paper, Béatrice Bourdin and Michel Fayol (1994; and subsequently 2000, 2002) explained, in terms of cognitive load, that low-level processes of written language production are more difficult than those involved in oral language production due to the cognitive costs of grapho-motoric and orthographic behavior.

Their results were obtained with mother-tongue speakers of French, a language with a rather little transparent orthography. The main aim of the present paper is to replicate the basic findings of Bourdin and Fayol in German speakers. These replications include the assessment of individual memory span in the written as opposed to the oral mode, as well as in elementary school children as opposed to university students. Moreover, we examine developmental changes of the cognitive costs of writing during elementary school.

The French results replicate quite well for German participants. Elementary school children, but not university students show worse memory span performance in writing, compared to oral recall. This effect holds for second graders as well as for forth graders, although an age-related overall increase in the number of correctly recalled items occurs in either recall mode.

Conclusions from these date relate, among other things, to the fact that until the end of primary school knowledge or cognitive abilities will not be assessed equally well in speaking and writing.

Bourdin and Fayol’s findings: A summary

Theoretical approach

By posing the question: „Is written language production more difficult than oral language production?” (Bourdin & Fayol, 1994), the authors started to report on a line of research that examined the cognitive costs of so called low-level activities in writing, i.e. formulation (lexical access and sentence generation), and graphic transcription and execution, as opposed to high-level processes that involve planning, i.e. the generation, selection and ordering of ideas. They based their expectations on a general resource model of the cognitive system, assuming that there is a limited capacity of working memory, that high-level as well as low-level components of writing put load on working memory, and that with increasing load of one component, performance based on the other components would suffer due to a decrease of the remaining resources. Bourdin and Fayol further assumed that children, but not (literate) adults, lack to some degree automation of low-level processes in writing (particularly spelling and the motor activity of handwriting) which should impair the performance of higher-level activities. (See Alamargot & Chanquoy, 2001; Grabowski, 1996; Kellogg, 1999; McCutchen, 1995; for more general accounts on cognitive processes in writing and writing development.)

In order to specifically empirically address low-level processes of language production, they used a task which did not call for planning, idea generation, or sentence formulation, thus excluding high-level processes as far as possible. Rather, participants were acoustically provided with word lists of increasing length and were, after the presentation of each list, asked to recall the list completely and in its original order, either in the oral or in the written mode of language production. Although this task
resembles a simple word span test (see, for example, Daneman & Carpenter, 1980), Bourdin and Fayol did not determine individual memory spans, but analysed the proportions of correctly recalled words within each list length.

The key experiment

In their basic experiment (= Experiment 1), 24 university students were presented with lists from four to nine words long, 24 second graders and 24 fourth graders with lists from three to seven words long. There were always two specimens of each list length, one of which was to recall orally, while the other one was recalled in written form. Thus, participants constantly changed recall modes, while the list length steadily increased. Of course, mode order and assignment of list length specimens to recall modes were counterbalanced.

Results include the absence of a mode effect for university students, i.e. serial recall performance remained consistent across written and oral recall; a significant mode effect for both groups of school children who show better recall in the spoken mode as compared to the written mode; and a significant effect of the children’s age, according to which fourth graders do generally better than second graders on the serial recall task. These results indicate that in second graders as well as in fourth graders, writing, which is not yet fully automated, involves higher cognitive load than speaking, thus reducing working memory capacity available for the serial recall task.

Further experiments

While Experiment 1 can be considered the key experiment of the Bourdin and Fayol (1994) research, the researchers added a series of experiments to support their interpretation against several possible objections. From these experiments, we learn that it is not the faster execution speed of speech as opposed to writing that facilitates the memory recall task as such, thus leading to better oral performance (Experiment 2); that children in a dictation mode, where the transcription of the recalled items was done by the experimenter, performed like in the oral mode (Experiment 3a), whereas performance of adults decreased in a mode of unpractised writing (Experiment 3b), suggesting that a lack of automation in handwriting could indeed be responsible for the overall result pattern. While all of the results discussed so far should, in principle, apply to the members of all linguistic communities (at least restricted to those who employ alphabetic script), because their interpretations relate to general cognitive processes, the authors continue the series of experiments by considering some aspects that could also play a role but are specific to French and the used verbal materials, like the more or less regular spelling of words and their familiarity (Experiment 4).

In subsequent research, Bourdin and Fayol continued to address the issue of cognitive costs of graphic activity in writing. In Bourdin and Fayol (2000), the non-automated processes in children’s writing are considered a secondary task which is, within a dual-task paradigm, compared to other secondary tasks during serial recall of word lists of increasing length. Again, oral recall proved to be superior to writing in primary school ages. In terms of secondary task load, written performance compared to a condition in which a sound categorization task was to be handled in parallel to the oral recall of the word lists, which clearly consumed attentional resources from working memory’s central executive. Through the careful dissociation between motor, graphic and lexical aspects by appropriately designed secondary tasks beyond the very oral and written recall, the authors also showed that the cognitive load of a simple drawing task appeared to significantly decrease from second to fourth grade of primary school, which indicates that a mere lack of automation in graphomotoric skills can not be (solely) responsible for the apparent cognitive costs of written recall, that impair serial recall performance when compared to the oral mode.

Eventually, Bourdin and Fayol (2002) again compare the cognitive costs of oral and written language production. Here, however, they directed their interest to adults as well as to a more complex text production task. Although the overall research question of identifying the cognitive load of low-level processes in writing is maintained, the study goes beyond the scope of replications discussed in this paper.
Estimating Bourdin and Fayol’s research

There are several aspects according to which the above-reported line of research deserves particular estimation, as seen from the perspective of the present paper’s author who as well is interested in the functioning of the diverse sub-processes of oral and written language production in terms of working memory load.

First, language production is a process the complexity of which is often beyond systematic experimental control. Research on language comprehension processes is much more advanced, at least historically as well as in terms of research output, because the materials are easier to control (Bock, 1996) and measures of understanding are easier to obtain in form of closed-class variables (which are, in turn, easier to analyse). Consider, for example, the Handbook of Psycholinguistics (Gernsbacher, 1994) that contains only one chapter among 34 chapters that is explicitly devoted to language production (and still restricted to grammatical aspects)! On the contrary, Bourdin and Fayol found appropriate tasks and experimental variations to decompose the language production process such that the interesting aspects remain under full experimental control. Language production involves hierarchically ordered sub-processes, from planning to motor execution, which can, therefore, only be controlled from the top downwards or from the bottom upwards. By employing a serial recall task, Bourdin and Fayol nicely eliminate any variation due to planning processes, syntactic processes, and processes of lexical selection. The remaining processes and their explanation, however, still appear difficult and complicated enough to be pursued for years.

Second, after an early period of investigation in which speaking and writing have been mainly compared with respect to their products (Blass & Siegman, 1975; Hidi & Hildyard, 1983; Horowitz & Newman; Portnoy, 1973) or their temporal economy (Gould, 1978), both modes of language production spent their scientific lives mostly separately and moved in different theoretical circles (Alamargot & Chanquoy, 2001). Particularly with respect to educational aspects, however, it would be helpful as well to directly compare the functioning of both verbal modes on all involved processing levels. For example, it is far from clear whether speaking and writing offer equivalent possibilities for students to express their cognitive achievements. Here, Bourdin and Fayol offer a theoretically uniform approach to the comparison of oral and written language production on low-level processes; similar proposals for high-level processes came from Herrmann and Grabowski (1995).

Reading Bourdin and Fayol’s results from another educational perspective, they point to a broadly neglected problem, namely the role of language production for diagnostic processes in primary school. If, as the above-mentioned results show, speaking and writing differ with respect to the loads they impose on working memory resources and, therefore, lead to significantly different performances on tasks as simple as serial recall of word lists, there is no reason to assume that both verbal modes would provide equivalent indications of more complexe cognitive processes or knowledge in long-term memory, where much more processes are involved that could differ between speaking and writing in relevant ways.

Together, the obtained results appear relevant enough to show whether they would also emerge in a population different from French primary school children. If the authors’ interpretation is correct, the findings should allow for successful replication in the young speakers of another language, because the observed differences between the oral and the written mode of recall relate to general cognitive principles.

Experiment A: Replication of the key experiment

The first experiment is aimed at the hypothesis of Bourdin and Fayol (1994, Experiment 1), according to which „in children, but not in adults, there is a greater load on working memory in the written mode than in the oral mode, due to the fact that the latter is mastered better than the former“ (p. 597). We used a serial recall task quite similar to the French original. With respect to the mode of analysis, however, we will not adopt the original procedure for three reasons. First, the way Bourdin and Fayol (1994) obtain their results from a a two-step analysis of proportions of recalled words appears, in statistical respect, neither smooth nor convincing. Second, it does not connect to memory span measures, although the applied method and task is pretty much the same. Third, it takes a standardized measure like memory span (or more complexe spans: cf. Ransdell & Levy, 1999) to compare results on serial recall tasks across materials and procedures. Therefore, we will describe
the obtained results in terms of simple word spans, in the sense of Daneman and Carpenter (1980), as the number of words that, after hearing them once, can be recalled in the correct order.

Method

Participants

Two samples of participants took the test. N = 24 children were in third grade of a German primary school (13 boys, 11 girls; mean age: 9.6 years, ranging from 9.0 to 10.3 years). N = 40 teacher students were from the Heidelberg University of Education (7 male and 33 female; mean age 23 years). With respect to age, these German third graders correspond to the fourth graders in the Bourdin and Fayol study; their mean age was 9.7 years.

Materials

252 frequent bi-syllabic German nouns, 5 to 8 letters in length, were selected from CELEX database. They were divided into two lists that contained word sets from three to nine words (= set size). There were three sets of each size in both lists (2 x 3 (3 + 4 + 5 + 6 + 7 + 8 + 9) = 252). For the children sample, the sets three to seven words long were used; for the students sample, the sets four to nine words long were used. Thus, the words in the set sizes four to seven were the same for both samples in both lists. Words were randomly assigned to the lists and sets.

Procedure

Participants were tested individually in a separate room of their school or university, respectively. After instruction to the task pattern, there was a practice trial with three sets of two words each to become familiar with the recall task. All sets were orally presented by the experimenter, at a rate of one word per second. Participants were instructed to start recall right after they heard the complete set. In the oral mode, the experimenter directly compared the participants’ recall with a prepared control list and indicated correct or false responses. In the written mode, a booklet had been prepared containing subsequent pages for each set of words.

The sets from one list were taken for oral recall, the sets from the other list for written recall. Half of the participants in each sample began with the oral condition, the other half with the written condition. Each condition started with the shortest set size (three words for children, four words for students) and continued until none of the three sets of a given size was successfully recalled. The assignment of the lists to the conditions, as well as to their order in the experiment, was systematically counterbalanced. Thus, for both samples, there were four groups who started either with sets from list A or from list B, and either with the oral condition or with the written condition.

This procedure differs in two respects from the French original. First, Bourdin and Fayol permanently switched between oral and written recall on each set length. Thus, their participants were to prepare themselves for another recall mode after every single recall performance. Particularly for the children’s sample, it could have additional costs to remember which kind of recall is expected for the next trial. Therefore, we ran a whole mode condition, from the shortest to the longest set size, and then repeated the entire procedure in the other verbal mode. Second, Bourdin and Fayol tested recall performance only by one set of words of a given length within each condition. This appears too susceptible to attentional distractions. Instead, we used three sets of a given size, which also facilitates the calculation of a span measure according to existing standards (see below).

Analysis

The word span score that is assigned to an individual’s performance in one of the two conditions is the highest size, at which all three sets have been correctly recalled. If one of the three sets of the next size has also been correctly recalled, 0.5 is added to the score. If two of the three sets of the next size have been correctly recalled, 1.0 is added to the score. However, correctly recalled sets within a set size are only considered for the score if all sets of the previous size have been recalled successfully.
Results and discussion

There were no effects due to the assignment of lists to conditions or to the order of conditions within a session; thus all of these control conditions were grouped together. The word span means of the four experimental conditions were as follows (standard errors given in parentheses) (Fig. 1): children/oral 3.56 (.13); children/written 3.13 (.11); adults/oral 5.06 (.10); adults/written 5.11 (.12). An analysis of variance with „oral vs. written span“ as repeated factor and „children vs. adults“ as between factor revealed a significant main effect for the age groups (F(1, 62) = 152.5; p < .0001) as well as a significant interaction between recall mode and age group (F(1, 62) = 7.88; p < .01). When comparing oral and written spans pair-wise within the two groups, oral word span appears to be superior to written word span in children (T(23) = 3.99; p > .001), but not in adults (T(39) = –.43; p > .5).

![Fig. 1: Mean oral and written word span by age; experiment A.](image)

The obtained findings perfectly equal those obtained from French participants in the Bourdin and Fayol (1994) study. Although the procedure was somewhat aggravated, compared to the French original, German primary school children obviously experience similar loads on working memory as French children of that age do, due to there not yet fully automated writing skills.

Experiment B: Replication of the developmental effect across primary school grades

In addition to the key finding of increased cognitive load of low-level writing processes in children, Bourdin and Fayol (1994, Experiment 1) also found that fourth graders performed better than second graders on both verbal modes of the serial recall task. Surprisingly, no interaction between age and mode appeared, which indicates that primary school children develop, from second to fourth class, some general cognitive ability, but that their automation of low-level writing skills does not substantially catch up to get any closer to oral performance. Would this picture also emerge in German primary school?

Method

Participants

Students from two classes from two different schools, all different form Experiment A, performed word span tests in the oral and written mode. N = 16 children were in second grade (6 boys, 10 girls; mean age: 7;10 years, ranging from 6;11 to 8;5 years). N = 20 children were in fourth grade (8 boys, 12 girls; mean age: 10;0 years, ranging from 9;5 to 10;10 years).

Materials

150 frequent bi-syllabic German words, 5 or 6 letters in length, were selected from CELEX database. Unlike experiment A, words were not restricted to nouns but could pertain to different parts of speech.
They were divided into two lists that contained word sets from three to seven words (= set size), three sets per size in each list \((2 \times 3 (3 + 4 + 5 + 6 + 7) = 150)\). Words were randomly assigned to the lists and sets, except for cases in which semantic relations between words within the same set were removed.

**Procedure and analysis**

Students were tested individually in a separate room of their school. All features of the procedure and the data analysis were the same as in Experiment A, including counterbalance of mode order and list assignments to modes.

**Results and discussion**

There were no effects due to the assignment of lists to conditions or to the order of conditions within a session; thus all control conditions were grouped together. The word span means of the four experimental conditions were as follows (standard errors given in parentheses) (Fig. 2): second-graders/oral 3.81 (.15); second-graders written 2.84 (.13); fourth-graders/oral 4.30 (.11); fourth-graders/written 3.63 (.10).

![Fig. 2: Mean oral and written word span by age; experiment B.](image)

An analysis of variance with „oral vs. written span“ as repeated factor and „second-graders vs. fourth-graders“ as between factor revealed a significant main effect for the age groups \(F(1, 34) = 19.1; p < .001\); overall performance was better in fourth-graders than in second graders. Further, a significant main effect of the recall mode emerged \(F (1, 34) = 87.1; p < .001\); oral performance was consistently better than written performance.

Like in the French study, there is no significant interaction between recall mode and age. This means that there is a linear increase in both oral and written recall from second to fourth grade, but that there is no substantial decrease from second to fourth grade regarding the difference between oral and written performance. Thus, until the end of primary school, low-level processes of writing still put so much load on working memory resources, that it appears far from equivalent whether knowledge or cognitive abilities are assessed through oral or written language production.

**Conclusions and implications**

The data of the reported experiments clearly support the assumption brought forward by Bourdin and Fayol (1994) that for primary school children, written language production is more difficult than oral language production and that the involved difficulty already appears on low-level processes that
include orthographic and grapho-motoric behavior. Although this hypothesis relates to cognitive processes that should generally apply to children of the species of human beings, it is nevertheless necessary to carry out cross-linguistic replications, when linguistic materials play an important role in the experimental task. The fact that exactly the same effects as in France emerged for German children, German adults and German word lists, can be considered a strong indication in favor of the given explanations.

With respect to educational implications, it appears particularly interesting to see that until fourth grade in elementary school, children perform worse on the same task when it is given in the written mode. Since the reported experiments involved only very simple tasks, it is quite likely that for more complex tasks which involve planning and recall from long-term memory, the written mode will pose even stronger restrictions on children’s cognitive resources. Thus, it could be advisable to rely not too much on written performance when the maximum abilities of children shall be assessed. Rather, comparing the oral and written performance on the same task could serve for the teacher as an indication of the progress that a student made in his or her writing skills; the more the written performance on a task equals the oral performance on the same task, the more automated are this student’s low-level processes of writing, which is, in turn, an indication of successful writing instruction in primary school!

In general, research on effects of different verbal modes of language production calls attention to the necessity of becoming more sensitive for language-related processes in diagnostic assessments. Measures like working memory capacity or word span can hardly be considered a unique characteristic and a stable feature of a person, when they change depending on the verbal mode of assessment.

There are also questions that remain unanswered in the above-mentioned experiments but wait for further clarification. Thus, measuring word span is a task that involves recall from working memory. Here, children do better in the oral than in the written mode, while adults show no difference between the verbal modes. With recall of knowledge from long-term memory, however, a writing superiority effect has been frequently observed for adults (Grabowski, 1999, in press). How do these effects combine? Moreover, particularly writing subdivides, in a technical world, into handwriting, typing, or operating modern input devices. What about effects between different ways of writing within the written mode as such? There is preliminary evidence that having adults to use either well-practised or unfamiliar writing techniques does influence tasks that depend on working memory capacity, but does not affect recall from long-term memory (Grabowski, 2003). Similarly, we will still have to find out whether starting to write by typing on a keyboard would facilitate the development of certain writing skills in primary school children.

Acknowledgements

I am grateful to the university students who carried out the experiments on the spot: Kathrin Bothe, Michaela Görlinger, Meike Jochemczyk, Maria Miller, Veronika Neuß, and Johanna Ritter.

References


