Writing

Gert Rijlaarsdam*, Huub Van den Bergh*/**, Michel Couzijn*, Tanja Janssen*, Martine Braaksma*, Marion Tillema**, Elke Van Steendam*** & Mariet Raedts***

*University of Amsterdam, the Netherlands, **Utrecht University, the Netherlands, **University of Brussels, Belgium

Introduction

Learning to write and using writing for communication and learning, is not a natural activity, such as learning to speak. It requires a great deal of schooling. The kind of schooling teachers offer novices or more advanced writers, has changed over time.

The history of writing education starts in ancient Greece at about 500 B.C, where writing was a part of both rhetoric education and of a more elementary schooling for clerks and other craftsmen who needed a certain technical writing ability to record information for trade and administrative purposes (Murphy, 2001). It was not until the installation of the public school system in the 19th century that writing education started to gain mass. Utensils such as blackboard and chalk, stylus, slate, pencil or pen then gained popularity in classrooms. Until the second half of the 20th century, the purpose of writing instruction was mainly to teach mechanics and conventions: handwriting, sentence construction (grammar), spelling and punctuation. In the past half-century, writing teachers began to pay more attention to text, content, style, and creativity.

Another change that occurred in the past 50 years is the transition from product-oriented to process-oriented writing instruction, stimulated by researchers such as Britton, Moffett, Emig and Graves in the 1960-70s. In addition, at the break of the 1980s cognitive psychologists such as Young, Hayes and Flower started to perceive writing as a ‘problem-solving’ activity. This led to the design and validation of writing process models, with the specific aim of applying the acquired insights as tools for teaching writing. The process approach allowed teachers to detect and diagnose strengths and needs in the different stages of writers' writing processes, such as planning, idea generation, formulation, or the revision stage. Process writing has become fully accepted in modern writing

education and engrained in many modern textbooks.

The 1990s and 2000s witnessed three more changes in writing education. The first is an orientation on the concept of 'genre': the notion that texts do not only comply with abstract linguistic criteria, but also with certain rather loosely defined content criteria, which allow for categorizations by purpose and audience.

Another approach that has gained popularity is the socio-cultural approach, which starts from the premise that choices made by the writer are governed by social background and cultural demands and habits. The implication for writing education is that students become aware of how their environment shapes their writing and their learning-to-write. They learn to accommodate their choices regarding content, style and voice to the environment in which the written text functions.

Lastly, the writing classroom has undergone a metamorphosis in the information age: computer-assisted writing has become the norm. Students use the computer as a text processor, as a source of information, and as a tool for communication. Moreover, software has been developed to aid writing instruction, in the form of step-by-step text production and feedback providing tools. For a more detailed discussion on technology in the (writing) classroom, see Moreno et al.'s chapter in this volume.

In this chapter we will present an overview of what we know about writing processes, instruction in learning to write, writing as a learning activity, the assessment of writing, and a selection of issues in writing research. We conclude with a short discussion of current and future issues. To prepare this last section, and to understand more fully in which a complex context writing has to be learned and used nowadays, we start with a exploration of the new contexts of writing.

New contexts for writing

The written world is changing rapidly, moving into a 'Print culture sans print' (Baron, 2006: p. 28). Developments in technology create new forms of written communication, such as wikis, blogs, personal websites, e-lists and e-fora, many of which are open for anyone to publish on .

The writer's world is becoming heterogeneous in culture and language. In the USA one out of four

school-aged children lives in a household where English is not the primarily spoken language. In Europe the situation is similar. Bilingualism has become the norm. Home language and language of instruction are different for many students. In addition, every student in secondary education in Europe must learn at least one 'foreign' language, English in most cases. Many Europeans cross national and linguistic borders for work and study, with Global English as their lingua franca. In short, language users experience a world which is linguistically, textually and culturally heterogeneous. In this introduction, we will explore the effects of technological developments and the increasing heterogeneity of language contexts on the concepts of writing and writing education.

Changing concepts of writing

Technology changes *the concept of text*. Students in primary and secondary education use multiple technologies to communicate with others in digital space. In schools, students are confronted with multimodal 'texts' as learning resources.

> "Frequently writing is now no longer the central mode of representation in learning materials – textbooks, Web-based resources, teacher-produced materials. Still (as well as moving) images are increasingly prominent as carriers of meaning. The digital media, rather than the (text) book, are more and more the site of appearance and distribution of learning resources, and writing is being displaced by image as the central mode for representation" (Bezemer & Kress, 2008, p. 166).

Bezemer and Kress (2008) also show that multiple sources are available during writing, in all kinds of digital forms. Writing is now 'just' one of the production systems within Document Design or Information Design (Schriver, 1997). It has become quite easy to construct documents which are made up of many media instead of text only (e.g., multimodality, multimedia). Such documents are created by means of multiple tools (e.g., computers, mobile phones) and are intended for multiple purposes and readers. A text is no longer predominantly a linear sequence of words, sentences, or paragraphs. A text is built from multiple structures, combining hierarchy, linearity, and recursiveness. In addition, an oral, informal style permeates written language (Pérez-Sabater, Peña-Martínez, Turney & Montero-Fleta, 2008); newly invented codes from conversational modes of writing, such as email and texting are nowadays present in school writing and in children's literature.

These technological developments imply a change in what it means to be a *literate person* (Nixon, 2003). As technologies change, users envision new ways of using them for literate acts. Literate persons possess strategic knowledge, including knowledge of what search strategies to use in order to acquire information in complex, changing contexts. They are critical thinkers and consumers of information and aware of the variety of meanings inherent to multiple media forms. Last but not least, literate persons are continuously learning and updating their knowledge of how to use technologies and how to communicate within them (Chatel, 2002).

Technology also affects *production systems*: writing is no longer produced primarily by handwriting, with pen and on paper, but done on (which are getting smaller) screens (e.g., mobile devices), with various forms of typing and keyboards. Other symbols are used beside letters. Producing written text can be realized even *without writing*. It is likely that speech recognition will replace typing or handwriting. Students with learning disabilities or other persons with disabilities already use speech recognition systems to produce written texts. MacArthur and Cavalier (2004) found that, for students with learning disabilities, essays dictated by means of speech recognition were of higher quality than handwritten essays. Another example of technology which supports persons with disabilities is the use of keyboards or writing tablets directed by direct brain connections (Mayo Clinic News, December 7, 2009).

Technology increases *access to and dissemination* of texts. Via the internet students have access to information from all over the world, which allows them to form broader conceptions of how the world functions. The internet also confronts them with a broader range of genres than they would experience otherwise. For example, they experience other ways of telling stories and providing valid argumentation, ways that are uncommon in their own cultures. In addition, they are confronted with a variety of versions of their languages. This holds for the national language(s) as well as the lingua franca for non-Anglophones: Global English (Crystal, 2003). All speakers of English as well as non-native English speakers have access to numerous variations of the English language. This variety of textual, socio-cultural, pragmalinguistic and linguistic input affects cultural identity, as Black (2009)

demonstrates how popular culture and technology converge to provide a context in which adolescent language learners are able to develop a powerful, transcultural identity.

Now that text production on computers is supported by available production software, we can observe an increased *speed of text production* in various stages of the writing process. Word production, drafting, revising, and sharing texts may benefit from the use of computers. *Words* (and in the near future, clauses) can be filled in automatically as soon as the author types the first two letters of a word. Attention to lower-level aspects of writing (e.g., spelling) is less required during typing. Young learners may therefore be less motivated to learn to spell. Studies on the effect of writing with computers versus writing with pen and paper, however, do not converge. On the basis of a meta-analysis of 32 experimental studies, Bangert-Drowns (1993) concluded that word processing improved the quantity and quality of writing, especially for poor writers while Berninger, Abbott, Augsburger and Garcia (2009), found that, children (2nd, 4th and 6th grade) consistently wrote longer essays with faster word production in 'pen and paper' conditions than in keyboard conditions.

The availability of digital sources often makes *drafting a first version* of a text easier, although Nuvoli (2000) showed that primary school students had more difficulty in drafting with the computer than with pen and paper. Copying and pasting (from) sources might facilitate the production of texts but requires, at the same time, more effort and higher abilities to check coherence of ideas in the text in the revision phase (Oliver, 2004).

Online and between- draft revision is less time-consuming in digital surroundings than with paper and pen writing. Daiute (1986) found that students corrected more errors autonomously when writing with a word-processor on a computer than when writing with pen and paper, but that they revised more when they received revision prompts. She concluded that external prompts should be built into word processors. Grejda and Hannafin (1992) also found a positive effect of word processing on the amount of revision, but not on the quality of the final text.

Sharing and co-authoring, be it online or offline, are easier in digital word processors. However, collaborative writing in general and in digital space in particular, requires extra skills.

Partners must deal with differences of collaborative writing groups, different styles of authoring, different roles in writing groups, and changes in the needs of collaborative writing over the process of the project. All these skills have to do with social interaction and cognitive aspects of co-authoring, and regulation of these processes. In a series of studies a team at Carnegie Mellon since the late 1980's (see for instance Neuwirth, Kaufer, Chandhok, & Morris, 1990) designs complementary devices in addition to word-processors to support effective co-authoring.

All in all, it seems that the availability of new technologies produces new constraints on writing processes and cognitive load, while it requires extra tools to produce text (MacArthur, 2006, p. 19).

Changes in writing education

The direction of technological advancements can cause a paradox. Creating something on screen which resembles 'text' has become easier, but selecting valid sources and monitoring and reviewing the quality of text to ensure coherence requires much more effort from the writer. The complexity of the writing process increases *because of* technology, and writing therefore requires more self-regulation. This provides opportunities for educational design to support document production with tools for planning, formulation, editing, revision and co-authoring (Neuwirth et al., 1990). Writers must invest time and effort into learning to use these tools. However, we do not yet know under which learning conditions and for which types of learners these tools are effective (MacArthur, 2005; 2006).

Technology is changing writing teachers' jobs, as computer-assisted assessment is developing and becoming accepted (Jorge-Botana, Leon, Olmos & Escudero, 2010). Teachers can arrange (peer) feedback by means of computer tools (Bouzidi & Jaillet, 2009; Cho & Schunn, 2007).Teachers can present writers with readers by means of audio tapes (Berner, Boswell & Kahan, 1996), podcasting (Davis & McGrail, 2009) or videos (Rijlaarsdam, Couzijn, Janssen, Braaksma, & Kieft, 2006). In addition, digital learning environments may support teachers, when they videotape 'models' to explore and compare all kinds of coping and mastery strategies, such as organizing content for argumentative

texts (Braaksma, Rijlaarsdam, Van den Bergh & Van Hout-Wolters, 2004), revision strategies (Van Steendam, Rijlaarsdam, Sercu & Van den Bergh, 2010) and strategies for writing synthesis texts (Raedts, Rijlaarsdam, Van Waes & Daems, 2007). Teachers may also use screen films of text processing as learning tools for discussion and analysis with students (Lindgren, 2004).

Concluding remarks: Dilemma's in future writing education

The importance of learning to write can hardly be overestimated in today's information society. Many consider learning to read and write to be at the core of the school curriculum, serving learning in all content areas. Literacy education takes up a quarter of the instruction time in schools for 7- to 11-year-olds (OECD, 2009). In secondary education, an average of 16% of instruction time is devoted to reading and writing. However, intended and actual time spent on writing may differ. Applebee and Langer (2006) found that in U.S. schools, two-thirds of the students in grade 8 write less than one hour per week for homework, and students in grade 12 report that they never or hardly ever have to write a paper of 3 pages or more. Nystrand, Gamoran, and Carbonaro (2001) report more or less the same amount of writing activities for ninth graders for English and social studies. Now that the world of texts is changing, and technology is available for many, new concepts for the writing curriculum must be generated and discussed, which create relevant learning opportunities without requiring much more instruction time. In this introduction, we pointed out the variation that exists for learners: variation of genres, of language as input and of language for production. Whether this variation supports or hinders the development of genre awareness, linguistic norms, and stylistic flexibility is still unresolved. We need to develop theories on the conditions under which variation can support development.

Another issue for theoretical consideration is the role of so-called 'lower level skills', such as handwriting, typing and spelling. Transcribing is a basic skill allowing higher order skills to develop. The flow of generating content could be hindered when handwriting or typing is too effortful. Therefore, teachers in primary school devote much time to consolidation and automatization of handwriting. The same should be done for typing. However, now that handwriting is less often

practiced outside school because of the availability of text processors, it is debatable whether even more time in school should be devoted to handwriting (and typing), considering that this practice in lower-level skills might be at the expense of the training of higher order activities.

A last issue is how students execute their writing process in various languages (Fitzgerald, 2006). As many students will have to learn to write in the language of instruction (which is for some students their first language (L1), for many others an additional language (L2)), they also have to learn to write in an international language, for an international audience. Communication becomes an intercultural endeavor. This raises the question under which conditions the learning of writing in L1 and in L2 should be stimulated; as parallel or serial processes?

Writing processes

In essence, writing is thinking (Kellogg, 1999). Writing a text demands *mindfulness* and *effortful engagement*: writers must monitor and evaluate how well thinking and writing is going. Kellogg (1999) presented data about cognitive effort in various tasks: writing, reading and playing chess. At randomly chosen moments during task execution, a tone was played and students were required to signal when they perceived the tone (e.g., by pressing a certain key). The reaction time indicates the amount of effort involved in the task: the longer it took participants to react, the more involved they were. The results were clear: for undergraduates writing required as much effort as selecting a move for expert chess players.

Writing is effortful *and* complex. The expression "Writing is a complex process" has become a cliché in discourse about writing. This cliché results from the relatively large body of research on the writing process, on the components of the writing process and the relations between those components. Major contributions were made by John Hayes and Linda Flower (both then at Carnegie Mellon, Chicago) and Carl Bereiter and Marlene Scardamalia (at OISE, Toronto). Hayes and Flower designed the well-known writing process model (1980, 1986), updated in 1996 (Hayes, 1996). Their approach was inspired by Newell and Simon's (1972) problem-solving model. Hayes and Flower applied this model to writing, and introduced the thinking-aloud methodology to investigate writing

processes, and the expert-novice paradigm to examine differences between novice and expert cognitive behavior in writing.

Bereiter and Scardamalia's (1987) work comes from experimentation and fits in a cognitive developmental paradigm. They concluded that mature and immature writers differ in the structure of the writing process. For that purpose they designed two different models of writing; the complex knowledge transforming model, where content and rhetoric situation interact, and the simpler knowledge telling model, a memory dump model, which leads to associative chains of content in texts.

For a thorough review of these and other models of writing we refer the reader to Alamargot and Chanquoy's (2001) monograph. Here, we limit our review to a handful of studies, mostly within the framework of the Hayes-and-Flower model. This model contains three components: (1) the task environment, all materials external to the writer with the task at hand and the text-produced-so-far as important elements, (2) the writer's long-term memory, with knowledge about the topic, audience, genres, task approaches, and (3) a set of cognitive operations, globally distinguished in three categories: planning activities (with goal setting, generating of ideas and structuring of ideas), formulating and transcribing activities, resulting in some materialized language (text-produced-so-far) and revising activities, containing reading already written text, evaluating, revising (intentional activity) and editing (automatically). A monitor controls the interplay between the cognitive activities. Within the Hayes-and-Flower framework, McCutchen (2006) provides a thorough overview of cognitive factors that affect the development of children's writing. We will first present some of these and other insights obtained from writing research in the dominant language of the schooling system (L1) and add some from writing in an additional language (L2).

A short overview of some insights from research

Constituting process variables to predict text quality

From the Hayes-and-Flower model (1980), Breetvelt and colleagues distilled the main subprocesses in think-aloud protocols, of 15-years-olds writing documented argumentative essays (Breetvelt, Van den Bergh & Rijlaarsdam, 1994). They found that a limited set of cognitive activities

proved to be sufficient to explain more than 80% of the variance of the quality of the resulting text if process time was taken into account. This means that for some of the activities, the higher their frequency during a certain phase in the writing process, the better the resulting text quality, whereas in another phase of the writing process the relation could be negative. These findings indicate that writers must vary the distribution of activities over the process dynamically.

Independent skills co-operate dynamically

Whitaker, Berninger, Johnston and Swanson (1994) studied the extent to which various components of writing are connected. Participants (4^{th} to 6^{th} grade) performed a sequence of tasks, each representing a component of the Hayes-and-Flower model (planning, translating, and reviewing). Whitaker et al. observed no significant relations between these three skills. This implies that writers must co-ordinate this set of skills. That such a co-ordination exists, is demonstrated by Rau and Sebrechts (1996) and Ransdell, Levy and Kellogg (2002). These studies indicate the compensatory power of writing processes: different task situations lead to different processes but may result in similar quality of text. Rau and Sebrechts implemented three planning conditions: no planning (immediate writing), silent planning (five minutes, no written planning) and outline planning (five minutes, planning with written notes). The results showed a clear distinction between the two planning conditions, indicating that during the writing phase writers behave differently according their planning condition, so that they compensate in such a manner that the resulting text quality is not affected by these different planning conditions. When writers can spend time on content planning, without much support to organize ideas (as in the silent condition), they increase the number of idea options which they have to choose from during writing. They pause longer, as an indication of content planning, and most of their revisions are content revisions. Writers who had the opportunity to organize their ideas, in the outline condition, generate *and* choose from options during the pre-writing phase, while the silent planners had to delay this process of options choice until the actual writing activity. Ransdell et al. (2002) showed via another methodology that writing processes are flexible managed to promote text quality. They found that when relatively minor demands were made on working memory writers

slowed down the writing process, to avoid a negative effect on text quality. However, a more demanding concurrent task (remembering six digits) reduced fluency by nearly 50% and also reliably decreased text quality and sentence length. Van Weijen (2009) found that within age groups, better writers show more variation within tasks and between tasks than weaker writers.

Dynamic interaction between processes

From think-aloud studies we learnt that types of cognitive activities can form functional units, in which one activity makes use of the information from a preceding activity; and that such a functional relation may change during the writing process (Van den Bergh & Rijlaarsdam, 2007). The activity of generating ideas, for instance, can be part of various combinations. For example, generating may follow an activity as formulating (strings as writing something-then generate an idea, then writing it down etc.) or reading the already written text (strings as reading a sentence or paragraph already written as a springboard for another moment of idea generation). The functional relation seems to change in the course of the process. For the combination generating and reading-already-written-text, it was observed that the functional relation changed over time. When writers rely on the already written text as a springboard for generating ideas in the first half of the process, it affected text quality negatively, while this sequence of behaviors turned out to be effective in the second half of the process. So it seems that the same cognitive activity can serve various functions depending on its interplay with other activities and the moment of operating in the system (Van den Bergh, Rijlaarsdam, Janssen, Braaksma, Van Weijen, & Tillema, 2009).

Writer characteristics influencing writing processes

Some studies attempt to explain writing process from writers' characteristics. Van der Hoeven (1997) for instance showed that a student's ability to apply a cognitive activity (planning, revising) affected the writing process and through that effect, it affected the quality of the resulting text positively. Van der Hoeven (1997) measured revision skills of 11-year-olds with revision tasks, writing processes and resulting text quality. She found that students relatively highly skilled in revisions, postponed the re-reading, evaluating, and revising activity in the writing process: they

revised relatively little in the beginning and more towards the end of the writing process. Interestingly, the revision skill scores, in itself, were negatively related to the quality of text. This implies that the skill itself is necessary but not sufficient; writers have to apply the skill in the writing process.

Galbraith (1996, 1999) investigated the relation between personality and writing to test the assumption that different writers perform better under different circumstances. Galbraith expected a difference between the writing processes of *high self-monitors*, who control their expressive behavior to present themselves in a pleasing way to others, and the writing process of *low self-monitors*, who express their affective state directly, without filtering. Undergraduate students (both high and low self-monitors) either wrote an essay (without making a plan on paper) or made notes in preparation for an essay (without actually writing the essay). Galbraith found that high self-monitors produced more new ideas when they made notes in the planning phase, while low self-monitors produced more ideas during writing. Thus, for discovery, writing is useful for some writers as a planning tool, whereas planning content before writing is useful for others.

Task variables that affect processes

In most writing process studies, findings are based on one single task within writers. Van Weijen (2009) showed how unreliable this may be. Her findings imply that such studies heavily overestimate the observed differences between writers. Fifty percent of the differences observed are not due to differences between individuals, but to variation within individuals, due to tasks. In Van Weijen's study twenty first-year university students wrote four 30-minute argumentative essays, with access to source materials (fact sheets, data, citations, clippings), under think-aloud conditions, in their L1. The only task variation was due to topic. From the analysis of cognitive activities, Van Weijen concluded that in general about 50% of the variance was due to within writer variance.

Writing in L2 affects writing process

Roca de Larios, Murphy and Marin (2002) provide a critical review of L2-writing process research. Only recently, two important shortcomings in L1-L2-process research were overcome: (1) in many L1-L2 comparative studies writers wrote just one text in L1 and one in L2; these studies

overestimate the differences as a result of the language, because no indices were available about the variation within a writer in one of the languages; (2) researchers related process indices to text quality in just a few studies; without establishing such relations, differences in processes between L1 and L2 cannot be interpreted in a functional framework. However, Van Weijen (2009; Van Weijen, Van den Bergh, Rijlaarsdam, & Sanders, 2008) implemented a design in which writers wrote four texts in their L1 and four in L2. Her study provides us with several important insights. In general, an L2-task, to a certain degree, inhibits the flexibility of the writing process. Writers do vary their distribution of activities in L2 less than in L1, and seem to accommodate less to the task at hand in L2 than in L1 as a result of the L2-load. The L2- process is not only more predictable for different tasks than the L1-process, it is also more uniform as far as the distribution of activities within one (and the same) task is concerned. Van Weijen's study also shows that L1 proficiency is a better predictor of L2-text quality than L2-proficiency (Van Weijen, Van den Bergh, Rijlaarsdam, & Sanders, 2009).

Typical L2-writing process feature: L1-use during L2-writing

A specific phenomenon when writing in L2 is the use of L1 during writing and thinking. One may expect that when writers experience a high cognitive load, they will tend to think more in L1 instead of L2. Whether use of L1 while thinking during the process inhibits the process or not, is subject of debate (for instance, Uzawa, 1996; Woodall, 2002). Van Weijen found that L1-use in verbal protocols is related to L1-writing proficiency: the better a student is in L1-writing, the less L1-use can be observed in their L2-writing. The amount of L1-use is not related to L2-proficiency (Van Weijen et al., 2009).

Concluding remarks

The more we know the more complex the writing process seems to be, and the less straightforward the relation between process and the teaching of writing becomes. Nevertheless, we will highlight four issues from this review that could serve as heuristics when designing writing instruction.

(1) Many paths lead to a good text. Writers compensate for less developed subskills. Variation in

instructional content and approaches is needed to build on strongly developed capacities, and to develop weaker skills or strategies. This might also hold for writing instruction in L2 that must be based on existing L1 writing competencies, which will vary in a group of students.

(2) Not all students exploit the relevant cognitive activities: they must become aware of their existence and effect and they must have opportunities to use them. Not all students launch the relevant activities at the appropriate time. Students must learn what is effective at what moment and must understand the weak elements in their patterns of writing activities.

(3) The *quality* of planning, generating, structuring, and revision skill(s) plays a positive role in writing processes and producing quality text. Teaching students these skills in productive and complex situations, that is, in context and not as separate skills, may promote writing growth.

(4) Students *must learn to guide and monitor their processes*. They must build task schemes (Hayes, 1996). This requires observation, awareness raising, practice and reflection: the development of meta-cognition.

Learning-to-write

Here we discuss insights gained from systematic reviews of writing development, writing instruction, and learning-to-write. The focus is on text composition: writing entire texts for communicative and/or expressive purposes.

We start by explaining how aspects of writing skill develop over time as induced by education and practice with special attention to the development and function of working memory (see also Swanson et al.'s chapter in Volume 1 for additional information on the relationship between memory and achievement). Finally, we hone in on writing instruction for struggling writers, self-regulation in learning-to-write and the interaction between learner characteristics and instruction.

A developmental perspective on learning to write

Students usually need 10 to 20 years to acquire a level of writing skill that prepares them adequately for the demanding writing tasks in their personal and professional lives (Kellogg, 2008). The development of writing skill progresses through three 'macro stages'. The first two are common

among most students and adults: *knowledge telling* and *knowledge transforming* (Bereiter & Scardamalia, 1987). A third, advanced stage, *knowledge crafting*, is only acquired by people who write for professional purposes (Table 1).

<<Insert here about table 1>>

Writing education usually starts by teaching children how to encode their words, ideas and speech into the graphic representation of a writing system and into well-formed sentences. Children devote their attention to the technical act of writing, to formulation and orthography. They rarely envision their text any further than one or two thoughts or sentences. Bereiter and Scardamalia (1987) coined this the stage of *knowledge telling*. The main limiting factor here at this point of development is *working memory*. Our active thinking is presumed to take place in a working memory with limited capacity. It involves the retrieval, integration and disposal of verbal and other information. New, difficult tasks take up most of the capacity of working memory. Alamargot and Fayol (2009) indicate that formulation and transcription are the main writing activities at this stage and that planning and revision are subsequently developed.

When sentence generation and handwriting have become more or less routine and take up less capacity, young writers increase their expertise by learning to critically shape their text, during or after writing. Their aim is to arrive at some instance of the text that best represents their ideas. In this *knowledge transforming* stage, writers learn to invest their mental resources (attention, memory, processing) in the negotiation between ideas and intentions on the one hand, and various formulations and text representations on the other. While doing so, they learn to juggle the three component processes of writing: *planning*, *formulation* and *reviewing*. Consequently, in the knowledge transforming stage writing has developed into a fairly complex affair, requiring a great deal of mental resources to organize ideas, to construct-and-compare text representations and to orchestrate the sub processes of writing.

Most people will not develop their writing beyond this *knowledge transforming* stage, because it serves them well for the writing tasks in their personal and professional lives. Professional writers

may enter the *knowledge crafting* stage by learning to maintain several representations of the text in their working memory: the author's representation of how the text should be, the representation of what the text written-thus-far actually means, and *an imagined reader's representation*: how potential readers might interpret the actual and intended text. The writer negotiates among these three kinds of mental representations to optimize the written communication.

To arrive at this 'knowledge crafting' stage, writers must have many mental resources at their disposal. This requirement can only be met if the more elementary subprocesses take up little conscious attention and processing capacity. Planning, formulating, and reviewing are each constrained by the limits of working memory in younger children compared to older children (McCutchen, 1996). Individual differences in writing ability and self-regulating ability in same-age groups also vary as a result of differences in working memory capacity (Ransdell & Levy, 1996) and mastery of handwriting and spelling (Graham & Harris, 2000).

Considering the necessity to off-load working memory to progress to higher stages of writing expertise, sufficient practice is essential. Yet, maturation also plays a vital role: the working memory capacity increases with age, allowing the pupil to progress to writing expertise levels that encompass more representations and perspectives (Alamargot & Fayol, 2009).

We should note that these three stages of writing development do not exclude each other: the 'knowledge telling' expertise is not lost after the 'knowledge transforming' stage has been reached. Professional writers who are experts in 'knowledge crafting' are still able to limit their writing to knowledge telling and transforming, given a particular writing task and setting.

What good teaching of writing entails: Meta-analyses of writing instruction research

In the past decades, several attempts have been made to integrate writing research in meta-analytic reviews. A meta-analysis provides an integral analysis, yielding a more robust effect size of the experimental variable generalized across its various implementations. The 'effect size' is the proportion of the variability (standard deviation, s.d.) within the group of participants due to the independent variable or treatment. An effect size of .20 is considered as a small effect, an effect of .50

as medium sized, and an effect of .80 as a large effect (Cohen, 1988).

Hillocks' meta-analysis (1986)

The first meta-analysis of research on writing instruction was conducted by Hillocks (1986). He focused on students in grade 3 to college. From 500 empirical studies he selected 100 studies that met his strict methodological criteria. Then he distinguished four categories of *instructional methods*, for which he calculated the mean experimental effect:

Presentation: the teacher explains qualities of 'good texts' or effective writing processes; students study examples and use them as models for writing (effect size: 0.02);

Natural Process: the teacher ensures that students acquire experience in (free) writing on topics and purposes that they choose themselves (effect size: 0.19 sd);

Environmental: the teacher organizes practice and integration of specific sub-skills, such as content generation, organization or reviewing, leading up to the composition of full text; students discuss and practice in groups and exchange feedback (effect size: 0.44 sd);

Individualized: the teacher ensures that students practice certain sub-skills individually and in isolation from the composition of a whole text (effect size: 0.17 sd).

The 'environmental' mode appeared clearly superior to the three other methods. The environmental mode presumes that the teacher has a good understanding of the components of the writing process, and integrates them in given assignments in which group work pays off. The teacher must set clear goals for the exercises, and present students with explicit criteria.

Furthermore, Hillocks analyzed the selected studies on 'focus of instruction': types of content or (pre)writing activities which teachers believe have a salutary effect on writing.

Grammar and mechanics: students follow explicit instruction and practice in these 'technical' aspects of sentence construction and formulation (effect size: -0.29);

Free Writing: students choose topics, purpose, and style and in the process discover 'their own voice' and ideas; the texts usually go ungraded (effect size: 0.16);

Models: students are instructed to study exemplary pieces of text, along with less adequate samples, and imitate the 'good texts' during practice (effect size: 0.22);

Scales: students use lists, consisting of specific criteria, checkmarks or questions that assist them in determining text quality (of their own or someone else's text); usually for the purpose of providing feedback with a view to revision (effect size: 0.36);
Sentence combining: students learn to combine simple sentences into more complex ones by using connectives such as 'because', 'but' and 'consequently' to signal their relationship explicitly (effect size: 0.36);
Inquiry: students analyze teacher-supplied or self-collected data, in order to answer one or more questions; the focus is on the ability to describe, explain, interpret, compare and summarize; pre-writing is usually group work (effect size: 0.56).

A focus on inquiry as a pre-writing activity yields the largest progress in writing quality. The teaching of grammar was shown to have a detrimental effect on text quality, probably because the focus of attention switches from text quality to grammatical correctness. In general, Hillocks found that the choice of focus of instruction explains variation in text quality quite well and has 'a significant impact on changing the quality of student writing' (op. cit., p. 217).

Graham and Perin's meta-analysis (2007)

Graham and Perin (2007a) concentrated on students of grades 4-12. Their meta-analysis drew on a body of 123 experimental and quasi-experimental studies and incorporated research from earlier meta-analyses, such as Hillocks' (1986) study. The authors re-conceptualized the treatments to avoid overlap. Effect sizes included in the analysis were weighted by a quality score of the study from which they originated. For instance, they were weighted by the number of participants in the study.

To answer the question which instructional practices improve the quality of adolescent students' writing the authors selected interventions for which research studies yielded at least four different effect sizes. Interventions were categorized by type of instruction (Table 2)

<<Insert about here table 2.>>

Graham and Perin found greatly varying effect sizes, i.e. actual influences on writing acquisition. Two cases of 'explicit teaching' are among the most effective interventions (> .60):
Strategy instruction (explanation and some training of planning, revising and/or editing strategies)

and *Summarization* (explicit instruction and practice in summarizing texts). Strategy instruction is most effective if it adheres to an optimized instructional Self-Regulated Strategy Development model (Graham & Perin, 2007a, p. 451). Two cases of 'scaffolding' also yielded strong effects: *Collaborative writing* (having students work together to plan, draft, and/or revise their compositions) and *Setting product goals* (assigning to students specific and concrete goals for the text to be written).

Moderate effects (.40 – .60) were found for *Word processing* (using a computer program during text composition) and *Sentence combining* (teaching students how to construct more complex sentences by combining two or more simple ones).

The *Process approach* was effective only for students in grades 7-12, and if teachers were professionally trained to offer their students process instruction. Graham and Perin found only small effects (< .40) for *Process instruction* by grade 4-6 teachers who had no training in this matter. Small effects were also found for *Prewriting* (a scaffolding activity focused on idea generation and organization, such as brainstorming); *Inquiry* (have students develop content and structure for a writing task by analyzing concrete data with an aim of answering a question) and *Study of models* (presenting students with text samples, with a view to emulate the perceived qualities and patterns). Special mention must be made of the clearly negative effect of a grammar-oriented intervention on writing, reaffirming Hillocks' conclusion.

The meta-analyses interpreted in a developmental perspective

How do the findings of Hillocks (1986) and Graham and Perin (2007a) match with Kellogg's (2008) developmental framework? From a developmental perspective, most of the students will be in the 'knowledge transforming' stage, with some elements of the 'knowledge crafting' in their proximal zone. During their high school years, they are mainly occupied with finding the right verbal or textual expression for their ideas. The more students discover that first versions are hardly ever final versions, the more they will be susceptible to reflection on product and process levels with the aim to improve their text. That is why it is important for students in the knowledge transforming stage to consciously

develop their capability to improve texts.

Hillocks (1986) showed that having students perform writing tasks that are within their zone of *actual* development (things that they can learn without any external assistance) such as free writing, does not yield strong results. Yet writing tasks that draw on skills in the zone of *proximal* development (things they can learn with some external assistance) proved to be the most successful. An example of this is the 'environmental' teaching approach, in which students comment on each others' texts in order to improve them. This finding matches Kellogg's *knowledge transforming* stage in that it focuses on writer-reader interaction with a view to both text improvement and learning-to-write. Students experience the writer's and the reader's role; acquire a sense of their own reader's expectations and learn to compare and select various options in formulating and organizing ideas (see also Slavin and Hopkins' chapter in Volume 1 for an overview of effective learning methods in which students help one another master academic content).

The most effective interventions in Graham and Perin's (2007a) meta-analysis all draw on *strategy instruction*. This is consistent with Kellogg's (2008) notion of limited working memory: idea generation, formulation and text production take up many of the limited cognitive resources. Strategy use helps (young) writers to develop beyond the *knowledge telling phase*; it takes away the burden of having to 'discover' ways to operate.

Writing instruction for struggling and language disabled writers

Both Rogers and Graham (2008) and Graham and Perin (2007b) focused on struggling writers and analyzed single-subject design writing interventions. In single subject design studies, participants serve as their own control. Instruction is lagged in various ways among students to allow the researcher to draw causal inferences about the effects of treatment. Both meta-analyses provide evidence that strategy instruction (e.g. in planning, drafting, and editing) is effective for struggling writers.

Findings from Joseph and Konrad's (2009) review study (9 studies with a total of 31 participants) indicate that students with intellectual disabilities benefit from strategy instruction as

well: strategy instruction helps these students improve the quality and quantity of their writing. Gersten and Baker (2001) and Baker, Gersten and Graham (2003) reviewed research on teaching students with learning disabilities (LD) to improve their writing of stories and essays, and demonstrated the effectiveness of a process orientation (teaching writing step-by-step), providing feedback, revision, and elaborated dialogue on the text as well as teaching about various text structures and their relationship to writing genres.

From a developmental perspective, these teaching strategies help students to make the transition to a 'knowledge transforming' stage, bridging a gap that proved too wide to bridge on their own. The instruction asks for attention to particular steps in the writing process, to text quality and organization, and to varying interpretations of their text that would have remained hidden from these students if they had only been instructed in 'good models' or 'general criteria'.

In Graham and Perin's (2007b) study the use of word processor programs as opposed to pen and paper was shown to be effective for struggling writers: 77% of treatment data points scored above the highest baseline data point (i.e. pen-and-paper writing). Rogers and Graham's (2008) review study confirms this finding.

In his overview study, Troia (2008) pointed out that most troublesome for students with LD are text production, planning, and revision. A curriculum for a diverse group of students should place these sub-skills in a greater process writing framework. The curriculum should emphasize strategic routines that offer students some grasp of their writing, which helps them become aware of their strengths and weaknesses. Sufficient time should be devoted to elementary skills such as handwriting, spelling and fluency. The writing tasks should be no less meaningful, communicative, diverse and challenging than with regular students.

Writing and learning to write: Regulation

Process models such as the one presented by Flower and Hayes (1981) show a complex of sub- and micro-processes that may run both parallel and in a recursive fashion during text composition. This puts high demands on cognitive resources, in particular on the regulative capacity

of the writer's cognitive system. Since writing instruction has become more process-oriented, the necessity to become aware of subprocesses and to learn to regulate them effectively has become more prominent (Galbraith & Rijlaarsdam, 1999). For a detailed discussion on the value of strategy instruction in writing we refer to the chapter of MacArthur et al. in this volume.

When the writing task is to some extent above the student's level of mastery – which is usually the case in writing education – the student is not only supposed to complete the task as a writing task but also to learn from it as a learning task. This is a *double challenge* to the cognitive system of the writer-as-a-learner: the writing task must result in a completed text with sufficient quality *and* a positive change in the cognitive capacity to write (Rijlaarsdam & Couzijn, 2000). In this perspective, possibilities have been studied to support the self-regulative competence of student writers to allow them to juggle the various cognitive demands of writing and learning more efficiently. Instruction that is supportive of writing strategies, such as the SRSD model (Graham, 2006; Harris, Graham, Mason, & Friedlander, 2007), may help students to acquire conscious control over the writing process and to regulate its course as they see fit. (See also Schunk's chapter in Volume 1 for the distinction between performance and learning.)

Torrance, Fidalgo and García (2007), for instance, investigated the teachability of self-regulation in writing expository texts. The researchers aimed to determine whether sixth grade students gain long-term benefits from instruction in self-regulatory strategies for planning and revising, with respect to writing processes and written products. Two groups of participants were compared: one group received cognitive strategy training by model observation and emulation; the other group followed a more traditional, assignment-based curriculum. The training resulted in a substantial increase in pre-planning and text quality. Effects were maintained on a delayed post-test and generalized across genres. The researchers concluded that even sixth grade writers are cognitively sufficiently well developed to benefit from appropriately delivered self-regulation training.

Glaser, Kessler and Brunstein (2009) examined the effects in grade four of a strategic story-writing training on story quality and writing competence. Students who received the writing strategies

training in conjunction with self-regulation procedures were compared with students who were taught the same writing strategies without self-regulation procedures. At posttest and follow-up assessments, students of the self-regulated writing strategies condition outperformed students of the writing strategies condition in specific story quality (content, vocabulary, coherence), holistic story quality and self-efficacy concerning writing.

Another approach to support regulation of learning can be found in Rijlaarsdam, Braaksma, Couzijn et al. (2008). They propose that the acquisition of complex writing skill can be induced by observation. "An important phase in learning to write, at all ages, is learning to write by (self-) observing, evaluating and improving relevant cognitive processes: writing processes, reading processes or communication processes between writers and readers" (p. 53). In a number of empirical studies this research group demonstrated that observation of peer writers and readers supports the observer's understanding of 'what works' in written communication. Observational learning also enables the observer to apply that understanding in his own writing (Holliway & McCutchen, 2004; Lumbelli & Paoletti, 2004). Referring to the *double challenge* analysis of learning-to-write assignments described above, observation may help the writing student to focus on the learning part of the assignment, rather than devoting the limited cognitive resources to text production. (See also Schunk's chapter in Volume 1 for observational learning.)

Zimmerman and Kitsantas (2007) present a model of self-regulation with three cyclical phases: forethought, performance control and self-observation. Self-regulatory processes during the forethought phase (goal setting and strategic planning) and sources of self-motivation (e.g. self-efficacy and outcome expectancy) prepare individuals to engage in writing. Self-control methods (e.g. self-instruction, mental imagery and attention focusing) and self-observation during performance phase provide input for self-reflection phase judgments and self-reactions (e.g. self-satisfaction and adaptive of defensive inferences). For a more detailed discussion of these aspects see Zimmerman and Labuhn's chapter in Volume 1.

Learning to write and learner characteristics

The meta-analyses reviewed establish average effect sizes for students receiving a specific form of instruction. Such main effects assume that effective interventions are valid for whole groups of participants, irrespective of differences in learners' aptitudes, attitudes and beliefs. Interactions between learner characteristics and interventions remain hidden. This is unfortunate, since some of these traits may play a vital role in effective instruction, such as the student's motivation to write or the personal interest in the subject (Boscolo, 2009). We will present some relevant 'aptitude-treatment interaction' (ATI) effects, partly based on a re-analysis of studies in the Graham and Perin (2007a) meta-analysis.

Saddler and Graham (2005) addressed interaction effects in their study about the effects of peer-assisted sentence-combining instruction on writing acquisition, taking into account the (already existing) ability of students in sentence writing. Results showed that poor writers in the experimental condition gained twice as much as more skilled writers. These large differential effects demonstrate the relevance of the ATI methodology (Cronbach & Snow, 1977).

Ferretti, MacArthur and Dowdy (2000) and Ferretti, Lewis and Andrews-Weckerly (2009) also took ATI-effects into account in their study about the effects of elaborated goal setting on persuasive writing. Fourth- and sixth-grade students with and without learning disabilities wrote persuasive essays under two different conditions: a 'general' goal condition in which students were asked to write a letter to persuade an audience to agree with their position, and a 'elaborated' goal condition where students were given the same general goal plus explicit subgoals based on the elements of argumentative discourse. Results showed ATI-effects of grade and disability status: students with learning disabilities included fewer elements than normally achieving students did at fourth grade, but not at sixth grade. To determine whether differences in intelligence could explain the results, the students with learning disabilities were divided into two groups (IQ higher or lower than 80). No main or interaction effects involving IQ were found for persuasiveness or the number of argumentative elements in the posttest. By including IQ in the analysis, the researchers could rule out an explanation for the effects found.

Braaksma, Rijlaarsdam and Van den Bergh (2002) investigated the effects of two types of observational learning in the domain of argumentative writing (reflection on ill-performing or *poor writing models* versus observation of *good writing models*) for students with different IQ levels. Results showed an effect of 'model-observer similarity' on the acquisition of argumentative writing: low IQ students profited more from reflection on ill-performing models, whereas high IQ students profited more from reflection on well-performing models.

Kieft, Rijlaarsdam, Galbraith, and Van den Bergh (2007) examined the benefits of adapting learning-to-write tasks to students with different writing strategies or styles. They developed instructions for learning to write literary reviews in two different versions: one adapted to a 'planning' writing strategy (think first, then write), the other to a 'revising' strategy ('write and then revise and improve'). A *match hypothesis* was experimentally tested and partly confirmed. Students with strong tendencies to either plan or revise profited both from writing instruction based on a planning strategy; while students with a low tendency to plan or revise profited more from instruction based on a revising strategy.

Whereas explicit grammar instruction is associated with negative effect sizes for typical students (Graham & Perin, 2007a; Hillocks 1986,), teaching grammar and usage had positive effects for struggling writers (Graham & Perin, 2007b; Rogers & Graham, 2008). The difference in findings can be explained by two factors: the type of writers who were included in the meta-analyses (typical writers vs. writers with LD) and the teaching method (traditional grammar vs. modeling of grammatical skill followed by student practice). Differences between typical writers and struggling writers were also found for setting product goals and instructional arrangements for collaborative writing and revision. These interventions proved to be less effective for struggling writers (Graham & Perin, 2007b). In addition, Rogers and Graham's (2008) review underlined the value of writing treatments not included in prior meta-analyses such as reinforcing students' writing productivity, teaching strategies for editing and for constructing paragraphs.

The research discussed in this section demonstrates that not all types of writing instruction are

equally effective for all types of students. It is worthwhile to take interaction effects into account, both in research and in the application into educational practice.

Concluding remarks

There is no panacea for all kinds of texts, all kinds of writing goals or problems, all levels of expertise, let alone for all kinds of writers (as our ATI section illustrates). Thoughtful writing teachers will always weigh the evidence and interpret it in the light of the kind of learning they want to establish with a particular group of students. There is no need to treat instructional approaches as competing. We would rather treat them as complementary and part of a rich and varied writing curriculum in which students meet several options to extend their writing expertise (Galbraith & Rijlaarsdam, 1999).

Smagorinsky (2006) and Chapman (2006) use the term 'social turn' for a transition of writing education research toward more context-sensitive and individualized types of instruction. The teaching of writing has become more complex because no single approach is best for all students; rather, the social and cultural contexts of classrooms suggest that instruction should take into account factors from learners' experiences and social practices and build on them, rather than expecting all students to respond identically to the same teaching. Hillocks (2006) states that writing instruction in secondary schools has not reached its potential, mainly because it is rigorously attached to assessment practices. He pleads for more open-ended, exploratory, inquiry-oriented writing activities.

Having said that, we can still provide a list of recommendations for teaching writing with adolescents in secondary education, based e.g. on the findings of Hillocks (1986), Graham and Perin (2007a) and Rijlaarsdam et al. (2008), provided that they exploit all features of the digital environment:

1. Have students develop *a sense of purpose* in their writing. Establish or have the students discover and discuss clear goals and criteria for what makes a 'good text' for this purpose, and for this audience. Consider the active exploration in the form of inquiry activities of textual models that demonstrate the implementation of these criteria.
2. Teach students to *gather, discuss, analyze and organize ideas* before drafting; stimulate them to sharpen their skills of inquiry and to use the results as a basis for writing; interactive talk and dialogue is an important device in this stage. Organize learning-to-write activities in which students co-operate

in planning, content generation and organization (inquiry), drafting and revision based on feedback; build communities of learners/writers.

3. Focus students' attention on (strategies for) the *subskills* (planning, reading sources, formulating, revising etc.) within the writing process and to its *recursive* and *regulative* character, e.g. by having them observe and evaluate coping and/or mastery models and provide emulation activities.

Writing-to-learn

Writing-to-learn refers to writing activities aimed at increasing students' learning in content areas. The premise is that writing is not just a way of communicating or displaying what has been learned, but also a tool for acquiring content knowledge, developing understanding, and improving thinking skills. This 'learning through writing' can be applied in different subject areas – ranging from science to literature, and from biology to history – and at various educational levels (primary, secondary and tertiary education).

In this section we will first sketch the history of the writing-to-learn movement. Next, we attempt to clarify some of the cognitive mechanisms underlying writing-to-learn. Finally, we present some components of successful pedagogical practices of teaching writing in content area subjects.

A brief history of writing-to-learn

Writing-to-learn is also known as writing across the curriculum (WAC) or writing in the disciplines (WID). In the United States, the WAC-movement originated in university-level composition courses, taught in departments of English. These courses were required of almost all first-year university students, and traditionally encompassed literary analysis and skills-based writing instruction. In the late 1970s composition teachers began to professionalize their teaching. A source of inspiration was the work of the British educational theorist and reformer James Britton and his colleagues, who promoted expressive writing and argued for a closer integration of writing with education in all subject areas (Britton, Burgess, Martin, McLeod, & Rosen, 1975).

Early advocates of writing-to-learn in the United States were James Moffett, Peter Elbow and Janet Emig, among others. In her essay "Writing as a Mode of Learning", Emig (1977) argued that writing is as a unique mode of learning, since some of its underlying strategies and processes promote

learning in ways that other forms of communication cannot. ‘To write is to learn’, according to Emig.

The WID-movement (writing in the disciplines) emerged somewhat later than WAC, during the 1980s. In WID, the emphasis is less on students’ expressive writing-to-learn and more on introducing students into a disciplinary community, where they are to learn the discourse of that community and become familiar with its preferred genres, reasoning structures and types of arguments. Science education, for instance, does not just strive to educate students into the ‘laws of science’ but also to teach them to communicate about science and to participate in the societal debate on issues where science plays a role (Yore, Bisanz & Hand, 2003). In other words, a biology student must learn to write as a biologist, a history major as a historian, a literature student as a literary critic or literary theorist.

Writing-to-learn has a long tradition and still receives strong support from educators in the United States and elsewhere (Thaiss & Porter, 2010). Since 1975, some 2400 articles and books on WAC have been published, with some 240 empirical studies (Russell, Lea, Parker, Street & Donahue, 2009, p. 396). Despite this strong support, the effects of writing on learning have, as yet, not been sufficiently demonstrated. Ackerman (1993) reviewed 35 writing-to-learn studies and found that studies were inconclusive, with no robust learning gains. He suggested that the results show, at best, how complex the relationship between writing and learning is. Similarly, Ochsner and Fowler (2004) reviewed the literature on writing-to-learn, and concluded that student achievement as a result of writing-to-learn pedagogy has been minimal. In a meta-analysis of 48 writing-to-learn studies, Bangert-Drowns, Hurley and Wilkinson (2004) found that writing-to-learn had a small beneficial impact on students’ academic achievement. (For other reviews see; Applebee, 1984; Klein, 1999; Newell, 2006; Rivard, 1994; Tynjälä, 2001.) All in all, the results of writing-to-learn studies are contradictory and rather confusing.

Mechanisms underlying writing-to-learn

In his review of writing-to-learn studies, Klein (1999) took the discussion about the effects of writing on learning a step further by proposing four hypotheses that may explain how writing affects

learning, and by discussing the empirical support for each hypothesis.

Spontaneous Text Production (shaping at the point of utterance)

The first hypothesis claims that writers learn spontaneously as they produce text, making their tacit knowledge explicit. Writers 'shape their ideas at the point of utterance', in Britton's words. This hypothesis relies on the translating component of the Hayes and Flower model of the writing process (see section 2 in this chapter). Shaping at the point of utterance is expressive writing, based on discovery through free writing. The cognitive load is minimal; the emphasis is on encouraging students to keep on writing.

There is some evidence that writers generate new ideas in the phase of drafting, and that the most original ideas appear in the pre-writing phase, not in writing the final text. However, further evidence for this hypothesis is anecdotal. Klein's criticism of this hypothesis is that, although exploring one's own experiences and knowledge by writing can lead to the generation of ideas, it cannot lead to conceptual change or a revision of existing ideas.

Forward Search

The second hypothesis claims that writers learn when they reread what they have already written, and then generate additional knowledge by evaluating, selecting and organizing ideas in the previously written text. This is called 'forward search', because it suggests that writers' initial drafts stimulate their production of subsequent ideas. The hypothesis proposes that learning itself is situated in the reviewing component of the Flower and Hayes model of the writing process .

The evidence for the forward search hypothesis is weak. Expert-novice studies suggest that experts review initial drafts deeper than novices, but this does not necessarily mean that experts learn more. Galbraith (1992, 1996) refined the forward search hypothesis, by suggesting that discovery is a consequence of the writer's implicit disposition towards a topic and the emerging text. He argued that writing is a dialectic between the text written so far and the writer's disposition. Only when the disposition is coherent, will the text be coherent; otherwise, additional drafts must follow. Galbraith (2009) argued that the classical models of the writing process (such as the model of Flower and

Hayes) overemphasize explicit thinking processes in writing at the expense of more implicit processes. 'Discovery' in the classical models is seen merely as a side-effect of the explicit, goal-directed thinking processes. In contrast, Galbraith considers 'discovery' as an intrinsic part of the writing process. He proposed an alternative dual-process model that identifies two conflicting processes in writing: an explicit planning process, and an implicit text production process. This last, implicit process is not a passive output process but an active knowledge-constituting process in its own right.

Backward Search

This hypothesis claims that writers learn when they set goals for writing, and then generate knowledge to achieve these goals. Learning is a result of the planning component of the writing process model. The hypothesis implies that writers use a knowledge-transforming strategy, which involves elaborating a representation of the rhetorical problem and using the goals derived from this representation to guide the generation of ideas during writing (Scardamalia & Bereiter, 1986).

Studies on writing provide some support for this hypothesis. Research has shown that different goals lead to different texts, and that some students who transform their knowledge appear to use a knowledge transforming strategy (Klein, Boman & Prince, 2007), but the effects on learning have not been tested.

Genre

The fourth hypothesis claims that writers learn when they apply content knowledge to compose in a particular genre. Some genres, such as argumentation, explanation and analogy, require deep processing, including the construction of relations between prior knowledge and new knowledge. These cause-effect, if-then, problem-solution and other relations are critical to some disciplines such as science and mathematics. According to the genre hypothesis, the operations and forms of organization required by writing in a particular genre lead to equivalent operations upon content. As a result, knowledge is organized and ideas are linked.

The genre hypothesis has been researched more extensively than the three process hypotheses. Results are generally positive. It was found, for instance, that students with more genre knowledge

wrote better texts, and students who wrote better texts learned more (Klein & Kirkpatrick, 2010). Also, several studies showed that composing in elaborative genres (e.g., analytical essays or argumentative texts) significantly improved students' recall and comprehension of source texts, as opposed to restricted writing (e.g., short answers) (Langer & Applebee, 1987; Marshall, 1987; Newell & Winograd, 1995; Newell, 2006). In a recent meta-analysis of writing studies, Graham and Hebert (2010) showed that writing in response to a text being read (e.g., writing personal reactions, summaries, or notes) has beneficial effects on students' reading comprehension.

The four hypotheses differ in the level of expertise required for writing-to-learn. Spontaneous Text Production requires little knowledge about composing, since it involves informal, 'speechlike' forms of writing. Forward Search and Backward Search, on the other hand, require rather sophisticated writing strategies, such as rhetorical goal setting and reviewing. For the Genre hypothesis it is unclear how much knowledge of text structures students must have in order to be able to learn through writing.

Writing to learn and learner characteristics

The four hypotheses do not take differences between learners into account. These differences might influence when and how learning will take place while writing. Theories about differences in learning styles, writing styles, and personality have been applied to writing and writing instruction (Davidson-Shivers, Nowlin, & Lanouette, 2002; Jensen & DiTiberio, 1984; Torrance, Thomas & Robinson, 2000). Few studies focus on writing-to-learn (Galbraith, 1996; 1999; Kieft et al., 2007; Kieft, Rijlaarsdam & Van den Bergh, 2008; Klein, 2000; 2004). (See Furnham's chapter in Vol.2 for an overview on learning styles.)

As mentioned in section 2 in this chapter, Galbraith (1996, 1999) investigated the relation between personality and writing. He found that students who were high self-monitors tended to generate most of their ideas during note-taking *prior* to writing, in the planning phase, while low self-monitors generated most of their ideas *during* writing. Galbraith concluded that for some students writing by planning appears to be most useful, while for others writing by producing text works best

for generating new ideas. Klein (2000; 2004) studied students' writing processes while they were performing explanatory writing tasks. He examined which writing operations were related to higher gains in content learning in science. He found that, for university students, two factors predicted 86 % of cases in which students made explanatory gains; 'problem solving' and 'comparisons in text'. The results supported the notion of a meta-cognitive, problem solving model of writing-to-learn.

Besides writing style, other factors may account for learning effects from writing, such as students' motivation, achievement, and writing skill (Klein, Piacente-Cimini & Williams, 2007; Klein & Kirkpatrick, 2010). Some researchers suggest that students need to be skilled writers or high achievers, in order to be able to learn from content area writing (Bereiter & Scardamalia, 1987; Rivard, 2004). However, other studies have shown that using writing-to-learn strategies can be advantageous even for elementary school children (Boscolo & Mason, 2001; Boscolo & Carotti, 2003; Chambliss, Christenson & Parker, 2003; Klein, 2000; Purcell-Gates, Duke & Martineau, 2007).

Components of successful writing-to-learn practices

What are the components of effective pedagogical practices aimed at learning through writing? While much is known about the components of effective writing-to-communicate instruction (see the section on 'Learning to write', in this chapter), the ingredients of effective writing-to-learn programs are less clear. In their meta-analysis of 48 studies, Bangert-Drowns et al. (2004) found only a small average effect size for writing-to-learn programs (median .20). Treatment effects were found to be correlated with treatment length. Time per writing assignment also mattered: longer assignments were less effective. This might be due to motivational problems, especially in poor writers. Whether the treatment contained personal writing did not influence effect size, nor did the presence or absence of feedback. The contribution of meta-cognitive prompts, however, was significant. Treatments using prompts for students to evaluate their understanding, confusion and feelings in relation to the subject matter were more effective, although the effect sizes were small. (See McCormick & Dimmitt's chapter in Volume 1 for Metacognition.)

This meta-analysis suggests that a meta-cognitive position of the learner/writer could

contribute to learning. During writing students should be encouraged to reflect on their understanding of the topic of writing, their affective or motivational responses to the topic, and on what they are learning. Klein and Kirkpatrick (2010) developed a framework that explicitly focuses on the meta-cognitive aspects of writing-to-learn. According to this framework, students do more than just perform writing tasks, and more than just learn how to write. Instead, they learn how to use writing as a tool for learning in content areas. Based on this premise Klein and Kirkpatrick proposed a set of design principles teachers can use to create writing intensive content area units for science and social studies. Design principles are, for example: 'writing three or more times per week', 'reflecting on learning following writing sessions', 'pre-writing experiences to generate data' and 'modeling self-monitoring and self-reinforcement'.

For science learning, the use of the Science Writing Heuristic (SWH) developed by Keys, Hand, Prain and Collins (1999), was found to be effective across national settings, gender, ages, and science disciplines. The SWH approach comprises two templates: one for teachers and one for students. In the teacher template, the teacher uses a series of writing, reading, and discussion activities to support students' thinking about science topics. In the students' template, students are encouraged to investigate their own questions about an activity and to use scientific methods. In sharing their findings students are encouraged to use their own language. Results of experimental studies indicate that students who used the SWH performed better on conceptual science questions than students who did not (Hand, Wallace & Yang, 2004). Critical elements for the success of the approach are, according to Gunel, Hand and Prain (2007): the translation of scientific to everyday language, the engagement in rhetorical elements of text production, the re-representation of key concepts in different wording and writing for a real audience.

A recent, promising strand of writing-to-learn research focuses on different modes of representation as a means for promoting learning in science. Scientists communicate not only by verbal language, but also by combining verbal text with graphs, tables, diagrams, maps, drawings and mathematical expressions. When dealing with a particular science concept, students must be able to

move between these different modes of representation. Hand, Gunel and Ulu (2009) asked students to explain their understanding of a topic through writing and through using different modes of representation (text only, text plus math, text plus graph). The researchers found that students who were asked to embed mathematical representation into their texts performed better on test questions than students who just wrote or embedded graphs into their texts. According to Hand et al. (2009), teachers should encourage students to be able to understand the full range of modes used to represent a concept and how to use these modes as tools to solve problems.

Concluding remarks

Clearly, the slogan of the writing-to-learn-movement in the 1970s that 'to write is to learn' was too bold, and deserves nuancing. Writing does not automatically lead to learning. Whether students learn by writing appears to depend on many factors, including characteristics of the writing task, learner characteristics, teacher practices, and the context as a whole. However, recent research has provided some indications of elements of successful writing-to-learn programs, such as the use of meta-cognitive strategies during writing, the use of everyday language, the use of different modes of representing a concept, and engagement in rhetorical elements of text production.

Measuring writing

In education and research, assessment of writing is still problematic. A person's writing ability is not a construct which is directly accessible. It can only be approximated by making inferences from writing performance: written texts. These inferences are, however, easily complicated by measurement error, causing unreliable measurements. In this section, we discuss potential sources of measurement error. First, we discuss rater effects and present measures which have been suggested for reducing them. Second, we look into task effects. We round off by discussing whether all variation in the measurement of writing can be attributed to error, or whether variation is inherent to the nature of writing. This final paragraph, then, is about the construct of writing ability. (See for validity issues on a more general level chapters by Elliott & Kurz and by Royer & Randell both in this volume.)

Rater effects

Instability in writing performance between and within writers can sometimes be brought back to instability or variability within and among raters, instead of to "the actual performance of the ratee" (Scullen, Mount & Goff, 2000, as cited by Myford & Wolfe, 2003, p. 391). One of the pioneering works to date that illustrated rater variability is the Diederich, French and Carlton (1961) study. Diederich et al. had 53 untrained raters from different professions read, annotate and score 300 papers written by university freshmen with a 9-point holistic scale. The results showed a rather low interrater agreement: "94% of the papers received at least seven different grades, and no paper received less than five separate scores" (Wolfe, 1997, p. 251).

Several rater effects can be distinguished: (1) for example, when some raters avoid extremely high or low scores but instead opt for scores "at or near the scale midpoint" (Knoch, Read, & von Randow, 2007, p. 27), there is a central tendency effect. (2) Raters may also be biased towards some or a specific group of writers (Barrit, Stock, & Clark, 1986; Diederich, 1974 and Rigsby, 1987 as cited by Huot, 1990) or with regard to certain characteristics of the written product to be rated such as handwriting as compared to computer-processed writing. If this personal opinion or impression is reflected in scores, the term halo effect is used (Myford & Wolfe, 2003, 2004). We must note that differences in severity of raters are often labeled as a rating-problem. However, this does not seem adequate as ratings are at best represented at interval level, rendering means and variances (of raters) arbitrary.

To reduce and eliminate rater effects, several measures are recommended. The choice for any of the measures proposed depends on the assessment context, e.g. for high-stakes writing assessment, entry exams or writing assessment in a classroom, and on their feasibility.

Increasing the number of raters

When only one rater evaluates text quality, that rater may be inconsistent in his or her ratings (resulting in a low intra-rater reliability). Additionally, differences in ratings of one rater only partially represent differences in students' writing skills or proficiency. Usually less than half of the differences in ratings of one rater can be attributed to differences in text quality, from which differences in

students' writing proficiency can be inferred (Van den Bergh & Eiting, 1989; Schoonen, 2005). Consequently, assessment should not be based on judgments by a single rater (Gebril, 2009; Meuffels, 1994; Schoonen, 2005).

Schoonen (2005) and Gebril (2009) illustrate how the desired number of raters to reach a reliable decision can – in combination with the preferred number of writing tasks – be estimated through G-studies in Generalizability Theory (G-Theory). G-Theory allows for estimations of the magnitude of different sources of error simultaneously (Brennan, 1992; Cronbach, Gleser, Nanda & Rajaratnam, 1972). So, the relative contribution of task, rater and test-taker and/or interactions to the total variance is estimated. In a G-study the questions are asked of how well measures taken in one context generalize to another and to what extent a sample of measurements can be generalized to the population of tasks from which the sample was drawn. As a result, a G-study allows for an estimation of the number of tasks and raters needed to make a generalization to writing ability of an individual. From the Schoonen and Gebril studies it can be concluded that an absolute minimum of two raters is required to "[yield] a considerable amount of error reduction" (Gebril, 2009, p. 524). As the Schoonen (2005) study illustrates, the decision of how many raters should score student writing cannot be taken without taking the *number of tasks*, the *scoring procedure* and even *the trait* to be scored into consideration.

In cases where there are many texts to be rated, it is not always feasible or sometimes even counterproductive to have all texts rated by the same small panel of raters. In such cases, a design of overlapping raters can be used in which each rater rates only a small (random) sample of all texts (Kuhlemeier & Van den Bergh, 1998; Van den Bergh & Eiting, 1989). For each sample of texts, an overlap between raters is created. This model allows for testing whether the ratings are congeneric (represent the same underlying trait except for errors of measurement), tau-equivalent, or parallel.

Selecting experienced raters

Apart from the question how many raters should score student writing, it is also important to consider which type of rater is preferable to reduce error. Rater expertise, experience and/or

proficiency in the use of a scoring rubric have an impact on the importance raters attach to certain textual features and on the scores they assign, as is illustrated in a number of studies which look into the rating process of expert and/or novice raters (Eckes, 2008; Huot, 1990; Vaughan, 1991). From a majority of these studies it can be concluded that experienced raters such as language teachers, are preferable to inexperienced ones as inter- and intra-rater agreement is frequently higher among experts. Nevertheless, even if experienced raters assess writing, an absolute minimum of two raters is still necessary.

However, as Schoonen, Vergeer, and Eiting (1997) have illustrated, the decision for experts or novices may depend on the task and trait to be scored. When scoring language usage in relatively free writing tasks, expert readers with teaching experience are more reliable. When evaluating content and organization of such a task, lay and expert raters are equally reliable (Schoonen et al., 1997). For restrictive writing assignments, such as interlinear revision tasks (i.e. tasks in which writers are prompted with a finished, but imperfect text, which they are required to revise), lay raters are as reliable as expert raters in scoring content and usage.

Clearly delineated criteria and categories in a scoring rubric

Providing raters with clearly delineated scoring criteria increases the chance that different raters score written texts in the same way. The different steps and scores in a rating scale should not only be clearly defined but should also be accompanied by so-called *anchor* or *benchmark* examples (Kuhlemeier & Van den Bergh, 1998; Schoonen, 2005). These so-called stabilizing factors contribute to what Quellmalz (1982) defines as scale reliability: scales should be stable across sets of raters and rating occasions.

There are several possible scoring methods such as holistic scoring, analytic scoring or the use of either a primary-trait or a multiple-trait scoring method in which one or several traits or criteria salient to a specific writing task in a specific context are being assessed (Hamp-Lyons, 1995; Lloyd-Jones, 1977). Primary-trait and holistic scoring can be distinguished from analytic scoring methods as there is only one (global) rating per dimension. In analytic scoring methods raters assign multiple

scores to a writing product or to a specific dimension of that writing product. These scores may be combined into one final total score.

The choice for any of these scoring methods depends on the specific assessment context and purpose (e.g. classroom use or large-scale testing, L1 or L2 writers: Bacha, 2001; Hamp-Lyons, 1995) and occasionally on the specific type of task (Deremer, 1998; Schoonen, 2005).

Cushing Weigle (2002, p. 72) rightly remarks that "the literature is replete with arguments for or against various scale types for both L1 and L2 writers" both in terms of validity and reliability. There is research proving that holistic scoring tends to yield high interrater agreement (Barkaoui, 2007). However, this high degree of interrater reliability would predominantly be achieved by the raters' predominant focus on superficial aspects of student writing such as handwriting, length and mechanics (McColly, 1970; Grobe, 1981, as cited by Charney, 1984). Other studies argue that holistic raters are most influenced by content and/or organisation when scoring (Breland & Jones, 1984; Freedman, 1979 as cited by Charney, 1984 and by Huot, 1990). In that respect, some doubts may be raised about the validity of holistic scoring. However, such criticism is countered by Huot (1990) who, on the basis of an experimental study with novice and expert holistic raters, showed that no evidence was found "to conclude that holistic practices impede the ability of raters to read and assess the quality of student writing" (p. 227). In any case, holistic scoring should preferably be conducted with the aid of "explicit guidelines" to avoid "positive characteristics … [compensating] for or even [outshining] weaker characteristics" (Schoonen, 2005, p. 5).

Analytic scales are generally considered to be more reliable, because more scores are given to one text and as such these scores can "account for variable and varying aspects of test taker performance (and underlying proficiency) more sensitively" (East, 2009, p. 91). In this sense, analytic assessment improves reliability as there are more test items. Nevertheless, some doubts are raised about the validity of analytic scores as well. Schoonen (2005, p. 15) showed that "analytic scores collected with scoring guides are less generalizable [and thus less valid] than holistic scores collected with essay scales [of benchmark essays]". It should also be noted that raters may adjust their analytic

scores to match their holistic impression (Knoch et al., 2007; Meuffels, 1994; Myford & Wolfe, 2003, 2004).

Training

The use of a clear scoring rubric in itself is not sufficient without training to obtain inter- and intra-rater reliability (Barkaoui, 2007; Cushing Weigle, 1994, 2002; Gebril, 2009; Hamp-Lyons, 2007; Lumley, 2002; Schoonen, 2005). Training is usually organized by having norming sessions with the different raters involved in which the scoring rubric or rating scale steps are first explained and then practiced and discussed.

However, training may not always be entirely effective in eliminating rater variability as raters may still have a different focus, display distinct patterns of scoring behavior or differences in severity post training (Cushing Weigle, 1998; Deremer, 1998; Eckes, 2008; Kondo-Brown, 2002). Additionally, training sessions are often criticized for their focus on intra- and inter-rater reliability at the expense of validity (Cushing Weigle, 1994; Hamp-Lyons, 2007; Huot, 1990; Elder, Knoch, Barkhuizen & von Randow, 2005; Moss, 1994). For example, training is often said to distort the natural process of reading and evaluating as readers are expected to agree with each other on the final judgment of student writing and in doing so may have to ignore “valuable responses to a student text" (Huot, 1990, p. 254 referring to Hake, 1986; East, 2009).

Automated essay scoring

Automated scoring, which originated in the 1960s, is an attempt towards a more objective, quantifiable method for writing assessment, in which human error and consequent rater effects are eliminated (Cizek & Page, 2003). These Automated Writing Evaluation programmes (AWEs) work with Artificial Intelligence (AI), Natural Language Processing (NLP) and Latent Semantic Analysis (LSA cf. Deerwester et al., 1990; Landauer, Laham, & Foltz, 2000) and compare characteristics of written texts such as syntax, punctuation, vocabulary used and lexical complexity, or semantic content with features of other written texts previously scored by human raters or with other (standard) texts in the semantic field of the writing assignment. (For an overview, see Warschauer & Ware, 2006). These

applications are criticized precisely because of their dependency on previously scored writing by human raters.

In general, however, automated scoring is considered to be reliable as the same script is always scored in an identical manner and correlation coefficients between machine automated scores and human scores are high (e.g. Cizek & Page, 2003; Shermis, Burnstein, & Leacock, 2006; Warschauer & Ware, 2006). Nonetheless, in Cizek and Page's (2003) opinion, other measures of reliability should be used for AWEs, such as decision consistency measures and conditional standard errors of measurement. Additionally, agreement and consistency between scores from different scoring algorithms used with different AWEs should be investigated in more detail to draw conclusions about the reliability of automated scoring. Finally, more research in the field of AWE is necessary. This is illustrated by the fact that some studies show that AWE software programs can easily be tricked, as expert writers obtained relatively high scores on polished nonsensical essays (Warschauer & Ware, 2006).

Task effects

Schoonen (2005) reports a number of studies that demonstrate that "scores for different writing tasks often show moderate or low correlations". Lehmann (1990), for instance, reported correlations ranging from .10 to .46 between nine different tasks (among which were job applications, formal notes, descriptions, and argumentative essays). This is line with earlier findings by Van den Bergh (1988), who reports results in which task-related variance exceeds the proportion of variance related to either differences in abilities (between writers) or the proportion of variance related to raters. Van den Bergh (1988) therefore argued that writing assessments on the basis of single tasks are basically single-item-tests, which do not allow for generalizations about an individual's writing ability. According to Van Weijen (2009, p. 170) it is "both impossible and unadvisable to determine a writer's skill accurately based on the output of a single writing task".

The exact number of tasks to be included in the assessment of individual writing skill largely depends on the degree of variability of writing performance across tasks within the writer or the

(homogeneous) group of writers (Coffman, 1966), but also on the number of raters included (Lehmann, 1990; Schoonen, 2005; Van den Bergh, 1988; see also the section about G-theory, earlier in this chapter) and on the scoring method (Van den Bergh, 1988). Estimates of the required number of tasks in specific situations have ranged from four (Coffman, 1966) to fourteen (Van den Bergh, 1988), depending of the type of tasks, varying from functional tasks to essays.

In some educational settings, portfolio assessment has been adopted as a suitable assessment tool to prevent judgments based on single-task-performances (Belanoff & Elbow, 1986; Hamp-Lyons & Condon, 2000; Cushing Weigle, 2002). A portfolio contains a variety of writing samples, in different genres, written over a period of time, for different purposes and addressing different audiences. Oftentimes, it also includes drafts and corrected earlier versions of texts. As such, portfolio assessment should provide an indication not only of what a student can do at a specific moment, but also of what he or she has been able to learn over time. Nevertheless, as the multiple writing samples in portfolios are in fact often single samples of different moments, and as they are often rated by few raters, the advantages of portfolio assessment are probably overestimated.

The construct of writing ability: narrow and broad definitions

Writing ability is a construct which can be defined in either a narrow or a broad sense. In the broad definition, writing is a general underlying ability which is called upon by writing tasks across genres, communicative situations, cultures and rhetorical modes. The observation that performance may fluctuate between tasks (Schoonen, 2005; Van Weijen, 2009) has sometimes been a reason to abandon the assumption that measurements of writing performance allow for generalizations about writing skill in general. Some researchers therefore promoted indirect measurement of writing skills, usually by means of multiple choice questions (Culpepper & Ramsdale, 1982). Others defined multiple narrowly defined writing abilities. In the narrow sense, the construct refers to an ability to write in specific subareas of the domain of written communication. Often, different writing abilities are classified according to rhetorical mode (e.g., description, narration, argument) and genre (e.g., letter, essay, manual). Narrow construct definitions are often applied, for example, in intervention

studies (compare curriculum based writing assessment versus curriculum independent assessment). In education, too, the objective of assessment is often to establish writing ability in the narrow sense.

The empirical validity of narrow definitions of writing may, however, be questioned. Re-analysis of data from a Dutch national assessment at the end of primary education (Van den Bergh, Rijlaarsdam & Zwarts, 1992) showed that correlations between text quality scores for different types of tasks (i.e. tasks differing in terms of rhetorical mode) are not lower than correlations between the same types of task (i.e. tasks in the same rhetorical mode, differing only in terms of topic). To put it simply, an argumentative essay and a descriptive text are no more different from each other than two argumentative essays on different topics. This suggests that, at least at this proficiency level, narrow definitions of writing according to rhetorical mode are not valid, unless we adopt the view that texts differing only in terms of topic also represent different constructs. Task effects are, therefore, no basis for abandoning the broad construct definition, in which writing is a general underlying ability, called upon by writing tasks across genres, communicative situations, cultures and rhetorical modes.

Indeed, Van den Bergh (1989) investigated the empirical validity of the broad construct definition by means of a regression analysis, in which variations in text quality scores for five different tasks are explained by scores on theoretically related constructs, such as fluency, spelling, vocabulary, and stylistic competence, as measured by separate tests. The results indicate that the various writing tasks all call upon the identified set of underlying constructs, although regression weights depend on the scoring method. These constructs are therefore regarded as part of the broad construct of writing ability.

There are, however, examples of intra-individual variation, which cannot be fully attributed to task or rater effects. Verheyden (2010), for example, demonstrated that the predictive value of an individual's writing performance on one occasion had little predictive value for writing performance on a later occasion on a similar task. How can such variation be explained? As we have seen earlier in this chapter, writing ability is, for the majority of people, a construct in development. If the ability has not been mastered completely yet, the speed with which writing develops may not be the same for

different types of tasks. In addition, progress does not always happen in a linear fashion. As such, variation in different instances of assessment of individuals is to be expected. If correlations between two instances of writing assessment are low, this reflects reality. Different instances of writing performance may very well call upon one and the same ability, but the fact that development of this ability does not happen linearly causes unreliability between measurement instances.

Concluding remarks

To conclude, assessment of writing needs multiple task and multiple raters. Although this has been known for a long time, many conclusions of process and intervention studies are based on single task measurement and therefore overestimating results.

Research in Writing

In this section, we present and discuss issues in writing research. (For general issues in design and data analysis of educational research see the chapter by Kulikowich & Sedransk in Vol. 1.) Writing research has become an international endeavor, including a variety of disciplines -- from psychology (cognitive, educational) and sociology to linguistics and psycholinguistics. In earlier years, most research in writing was published in *Research in the Teaching of English,* published by the National Council of Teachers of English (USA) since 1967. Now specific journals on writing research are available (*Written Communication*, since1983; *Journal of Second Language Writing*, since 1991; *Assessing Writing*, since 1994, *Journal of Writing Research*, since 2008). There is an international book series (*Studies in Writing*, since 1996, with 21 volumes), and several handbooks have been published recently (Bazerman, 2007; Rijlaarsdam, Van den Bergh & Couzijn, 2004; MacArthur, Graham, & Fitzgerald, 2006; Beard, Myhill, Riley & Nystrand, 2009). Furthermore, the international writing research community meets regularly at conferences (EARLI Sig Writing, biannual, since 1988; Writing Research across Borders, 4th conference 2011; Second Language Writing Symposium, since 1998, from 2006 annually). The amount of research has increased and as such we know a lot more about writing and writing education today than we did twenty years ago. We now know better which questions need to be studied. Some of these are addressed next.

Studying writing processes online

For the study of writing processes, a distinction can be made between online and offline methods. With the progression of technology, online methods are developing. With online methods, information on processes is gathered *during* writing. Online methods include techniques such as think-aloud protocols, key-logging, dual-task methods and recently, eye-tracking.

Think-aloud protocols are widely used in writing-process studies since the publication of the seminal papers by Hayes and Flower (see section Writing processes this chapter). To gather think-aloud protocols, researchers have participants verbalize all (the) thoughts that occur to them during writing. Researchers do not intervene in the think-aloud process by, for example, asking writers to explain why they are choosing a specific process, or a specific strategy. Participants just have to verbalize what they think during writing (Ericsson & Simon, 1993). Typically, these think-aloud responses are transcribed and analyzed for cognitive activities. That is, each segment of a think-aloud protocol is coded in terms of acts of planning, generating, revising, and so on. The most common coding schemes are based on the work of Hayes and Flower (1980) (For examples of categories see Breetvelt et al., 1994.)

A coding scheme functions as a guide for segmentation as it defines the units to be distinguished. Note that this point of departure in coding has bearings on a theory of writing processes. It is (implicitly) assumed that different cognitive activities are carried out successively. As it is impossible to verbalize two activities at the same time, and each segment is coded as one cognitive activity, parallel processing is not the main focus in think-aloud studies.

Research has shown that think-aloud protocols provide very rich data on the writing process. However, the think-aloud method is not without drawbacks. First, thinking aloud during writing might influence the writing process itself. This reactivity effect can take the form of participants devoting more attention to some of the cognitive activities carried out than they would without having to think aloud. For instance, hearing oneself generating ideas, or hearing oneself read a part of the text might serve as a trigger for further generation of ideas (Galbraith, 1996). While thinking aloud, writers

attend to the writing task itself and simultaneously verbalize their thoughts with an unknown effect. Empirical evidence shows that pauses, episodes during which the writer does not verbalize any thoughts, are functionally related to other cognitive activities (Van den Bergh & Rijlaarsdam, 2007) and that the number and distribution of these pauses correlates with text quality (Breetvelt et al., 1994). Here it can be concluded that when short-term memory experiences an overload (Kellogg, 1988; 2001) not all cognitive activities are verbalized, and that a resulting think-aloud protocol may therefore be somewhat biased, or impoverished. This kind of reactivity, in which the primary process (writing) is changed as a result of the way in which it is measured, can pose a threat to the validity of think-aloud protocols. Some research on the influence of thinking-aloud on writing processes is available. Results indicate that writing under think-aloud conditions takes more time, but that neither syntactic complexity nor text length (number of words) are influenced by thinking aloud (Ransdell, 1993; Janssen, Van Waes & Van den Bergh, 1996). In addition, the understanding of the task does not change either (Leow & Morgan-Short, 2006), and neither does the number and type of revisions (Stratham & Hamp-Lyons, 1994). There is some evidence that thinking aloud influences the cognitive organization of information (Biehal & Chakravarti, 1989), although such an effect appears to be dependent of the task (Russo, Johnson & Stephans, 1989).

Second, thinking aloud might not reflect all underlying processes, especially automatised processes, which the writer is likely not to verbalize. Therefore, think-aloud protocols might present an incomplete and therefore biased picture of the cognitive processes involved in text production (Jacoby, Lindsay & Toth, 1992). This is supported by evidence that sudden corrections of typing errors are not mentioned in think-aloud protocols (Tirkkonen-Condit, 2005). In addition, Galbraith (1999) argues that text production and idea generation proceed, at least to a certain degree, unconsciously by means of automatic spreading activation (Anderson, 1983). Information from the writing prompt, the text produced so far and the writing process itself are seen as the main source for creation of ideas during text production. Hence, the mere verbalization of thoughts renders input to the activation of memory nodes, and helps or influences the generation of ideas.

Third, for a variety of tasks it has been shown that the quality of think-aloud protocols is related to verbal ability (chess: De Groot, 1946; mathematics: Pugalee, 2004; learning from text: Kucan & Beck, 1997). High verbal ability respondents produce richer, more fine-grained think-aloud protocols than low verbal ability respondents. For writing however, no data are available.

All in all, think-aloud procedures can produce rich, detailed information on the mental activities going on during writing. The method is, however, not without validity problems

Keystroke logging. Most text today is produced with a word processor. This has made it possible to gather information on writing processes by using keystroke-logging. Two recent keystroke logging programs are *Inputlog* (Van Waes & Leijten, 2006) and *Scriptlog* (Strömqvist, Holmqvist, Johansson, Karlsson & Wengelin, 2006). Keystroke-logging software captures all keystrokes and mouse movements and stores this information for later analysis. This enables the study of writing in real time.

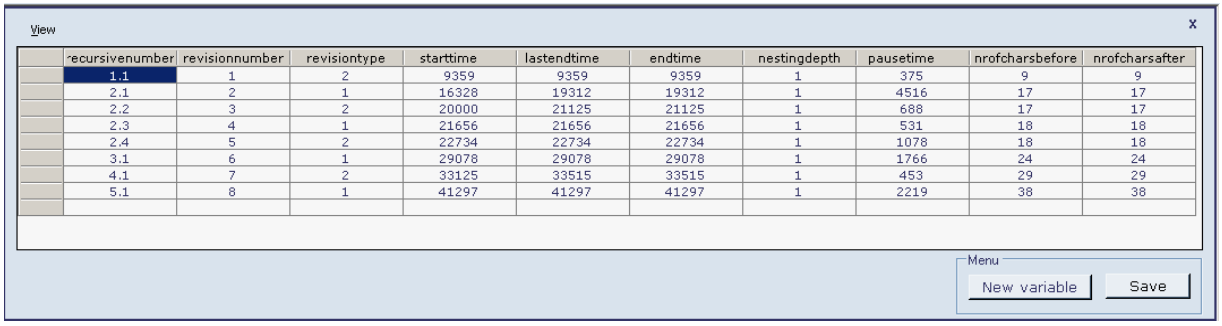
The rate at which writers produce text (i.e. the number of keys pressed without pauses) is interpreted as a measure of fluency (Uppstad & Solheim, 2006) and delay between consecutive keys is interpreted as cognitive effort (Butterworth, 1975) at that point in text production. It is assumed that the time which writers need reflects the amount of cognitive operations performed at that stage of text production. It has been shown, for instance, that pauses are not randomly distributed. Writers perform more cognitive operations between sentences than within sentences (Schilperoord, 1996), for instance. Keystroke logging programs provide an extremely valuable tool for writing process research. The precision with which production is recorded is unprecedented. Groenendijk, Janssen, Rijlaarsdam and Van den Bergh (2008) focused on the production of creative texts (poems) by 11th graders. They showed that differences in orchestration of production, pauses and different types of revision over the writing process collected via keystroke logging were related to the quality of the poems. For instance, production (i.e. the number of words produced) was positively related to text quality in the beginning of the writing process, whereas the number of revisions (at sentence level) was positively related to text quality at the end of the writing process.

A special feature of keystroke logging tools is that it allows for the study of revisions during writing in a more detailed way than is possible with think-alouds. Take for instance the revisions in the following sentence (Kolberg & Severinsson-Eklundh, 2003):

{Poeki}$_2$|^3[The cat]$_3$|^4[[h|^1]$_1${ }$_4$|$_5$ is b{l}$_6$ack wi[ht]$_7$]^7 yo{l}$_8$low spots.

That is: first *The cat h* is written, after which the character *h* is deleted. Then, *Poeki* is inserted and *The cat* is deleted (etc.). Several keystroke logging programs can generate such notations, but other representations are also possible as output. Table 3 displays Inputlog's (Van Waes & Leijten 2006) output for the same sentence. The first revision (1.1) concerns the 9th character (h) which was deleted (revision type 1). This deletion took place after 9359 ms and ended at the same time. There was a very short pause (375 ms) before this revision.

Table 3. Some information from the revision tool in the key-stroke logging program Inputlog


View

recursivenumber	revisionnumber	revisiontype	starttime	lastendtime	endtime	nestingdepth	pausetime	nrofcharsbefore	nrofcharsafter
1.1	1	2	9359	9359	9359	1	375	9	9
2.1	2	1	16328	19312	19312	1	4516	17	17
2.2	3	2	20000	21125	21125	1	688	17	17
2.3	4	1	21656	21656	21656	1	531	18	18
2.4	5	2	22734	22734	22734	1	1078	18	18
3.1	6	1	29078	29078	29078	1	1766	24	24
4.1	7	2	33125	33515	33515	1	453	29	29
5.1	8	1	41297	41297	41297	1	2219	38	38

Menu
New variable
Save

The data give an overview of revisions in real time and thus allow for a far more fine- grained analysis of revisions than was previously possible. However, every analysis, especially an analysis which is made by a computer program, uses definitions. The automaticity with which keystroke logging tools generate and categorize revisions is based on definitions. One must always approach these definitions critically. For instance, in the average time for sentence production, one of the units of analysis, depends on the definition of what constitutes a sentence. A sentence cannot always be characterized by: 'DOT<SPACE (or RETURN)> CAPITAL'. Hence, the resulting analysis will be biased according to the operating definition.

Keystroke logging clearly circumvents most of the validity issues of think-aloud methods. The

method cannot influence the writing process, nor are any coders needed before data can be analyzed and interpreted. However, the meaning of pauses in keystroke logging data can only be approached by means of reconstruction; the mere analysis of pauses does not provide insight into the cognitive operations carried out during these pauses. The occurrence of a pause may be indicative of a number of possible actions: a writer may have been thinking of a next sentence, she/he may have been rereading his text thus far, or reading some documentation and so on, or might have taken an off task pause. Differences in the length of pauses merely indicate the degree of cognitive effort needed at that moment in text production. It does not show what cognitive activity is going on in a writer's mind. Second, keystroke logging studies produce a lot of data. Without a specific theory and carefully designed experiments, it may be very hard to interpret all the information gathered by means of key-logging.

Dual and Triple task methods are based on the same assumption as pause analysis with keystroke logging tools: as resources are severely limited during writing, long pauses indicate a high level of cognitive processing (Baddeley & Andrade, 2000). During writing, all cognitive processes need to be coordinated as they compete for the same executive resources. Because these resources are limited, increases in resource use by the primary task will be reflected by less attention to a secondary task. In standard dual task techniques respondents have to react to auditory signals as fast as they can while executing the writing task (primary task). Differences in reaction time are interpreted in terms of differences in cognitive load during primary task (e.g. Kellogg, 1999). The weakness of this method is of course that it only provides an indication that writers are doing 'something' and not what they are doing.

In the *triple-task method* this weakness of the dual-task paradigm is circumvented. During writing (primary task), writers have to respond as quickly as possible to auditory probes (secondary task). In addition, they have to indicate what they are doing at that moment (tertiary task; Olive, Kellogg & Piolat, 2001; Piolat, Kellogg & Farioli, 2001). That is, after each auditory probe, subjects have to label their activity at that moment in one of the presented categories (e.g. planning, translating,

evaluating, revising). Garcia and Fidalgo (2008) showed that the procedures of the triple task technique can be used in classroom settings, even with students with learning disabilities. As such, this technique not only provides us with insight into what writers are doing, but also into how much cognitive effort is needed for this activity.

The question remains, however, whether reacting to auditory probes and/or labeling an ongoing activity influences the writing process. From the results of several studies it appears that a secondary task does not seem to influence the writing process (Piolat, Roussey, Olive, & Farioli, 1996). The influence of the tertiary task has been studied as well. Kellogg (2001) shows a large agreement between the labeling of the observed writing activity by the writer himself during writing and the labeling by trained raters afterwards. This indicates that writers are able to label their cognitive activities while writing. However, the labeling activity during writing (online) might disrupt the writing process itself. Therefore, one may consider asking writers to label their cognitive activities after the writing product is finished (offline). However, Levy, Marek and Lea (1995) concluded that there can be large differences between online labeling of activities and labeling afterwards during a stimulated recall. Piolat et al. (2001) concluded that the correspondence between online and offline labeling is *'not much better than random responding'* (p. 71). Hence, online labeling has to be preferred over offline labeling.

We must conclude that none of the methods for the online study of writing processes is perfect. All three online methods have shortcomings. A combination of methods is therefore recommended. For instance, Van Weijen (2009) used think-aloud techniques as well as a keystroke logging tool for her research on L1 and L2 writing (see also, Van Waes, Leijten & Van Weijen, 2009). This proved very useful for the segmentation of the think-aloud data. Wengelin, Leijten & Van Waes (2010) combined *eye-tracking* and keystroke logging. As such, they gathered information on the writing process, but also information on the reading process. Hence, (part of the) pauses can be interpreted in terms of reading activities. Cooke and Cuddihy (2005) combined think-aloud techniques with eye-tracking to validate the responses of participants thinking out loud. Results suggest that the

information from both sources overcomes limitations of either method on its own.

Data analysis of online data is complex. Researchers need to be aware of the strengths and weaknesses of various approaches. In the seventies and eighties, the analysis of think-aloud protocols was reduced to frequency analyses. That is, researchers counted the number of times writers carried out a cognitive activity (planning, generating, revising, etc.). Skilled and less skilled writers were compared in terms of the number of planning, generating and revising activities they employed. The consequence is that all time-related information is lost. Today, however, this time perspective seems to be essential: writing processes are now studied in real time, irrespective of the online method used. Think-aloud protocols are analyzed in time (e.g. Van den Bergh & Rijlaarsdam, 2001). Results show that the probability of occurrence of cognitive activities changes over time. During keystroke logging and multiple task techniques, time is an essential element of the method itself.

Intervention studies: Design issues

For intervention studies on writing to contribute to theory development, we need to consider at least four design ingredients. We point to the role of (1) pretest measurements (interaction between learner characteristics and intervention), (2) the dependent variables (implementing multiple writing tasks and process measures), (3) implementation measures and (4) efficient use of participants for replication (swapping panels).

Learner characteristics. The most frequently implemented research design is a *pretest-posttest design*. We propose to reconsider the concept of pretest in this design. From a measurement point of view, a congeneric pretest is impossible. If we teach students new things like how to write a letter of recommendation, they cannot be expected to perform well on tests before instruction. In a traditional pretest-posttest design we measure students' (lack of) skills both before and after a treatment. This may result in a pretest which appears to be unreliable. In such a case, one wonders why pretest results are taken into account. They can neither be used as a point of departure to show some kind of learning gain, nor as a covariate as they do not correlate with the posttest (mostly due to a lack of reliability). In such cases, a pretest of the construct under study is not very useful: related constructs might serve both

purposes better, such as fluency in writing. Therefore, and for theoretical reasons we recommend a theoretically driven pretest task in cases where something new is taught. The theoretical question then is whether we expect an interaction between a learner characteristic (e.g. prior knowledge or skill and type of writing style) and the intervention (see Braaksma et al., 2002; Kieft et al. 2008).

More texts in posttest measurements. To get an adequate estimate of the writing skill of a student, one would need multiple measures of writing (Coffman, 1966), preferably for both the pre- and posttest, but in any case for the posttest. Schoonen (2005), for instance, shows that the generalizability of the outcomes varies between .2 and .6 (depending on scoring procedures and task characteristics) if only one writing assignment is administered. Administering multiple tasks per participant, however, is not always feasible as the testing time could exceed the time devoted to instruction by far.

However, having students take only one assignment has far-reaching consequences: one can only generalize over the sample of students (if well sampled, which is hardly ever the case in intervention studies). That is, the research question is restricted to: if I were to give another sample of students the same treatments and the same assignment, what are the chances of finding a difference between the interventions? Or, to put it differently, there is at least one assignment which shows that there is an effect of the intervention. This is not the answer we want to produce. In fact, we want to generalize simultaneously to the population of participants and to a population of writing tasks (compare, Clark, 1973). Hence, differences between assignments (as well as differences between subjects) contribute to the error variance.

The result is that we might have been over-optimistic with regard to the estimation of effects of treatments. Studies in which several writing pedagogies are compared with multiple measures are scarce. Even in studies where multiple measurements are used, the differences between assignments are not taken into account. Such studies therefore present inflated effects. This is always the case if random factors (i.e. writing assignments) are treated as a fixed factor (Siemer & Joorman, 2003; Snijders & Bosker, 1999; Quené & Van den Bergh, 2008). As we want to generalize across

assignments and individuals, both should be treated as random factors. To this purpose, the mixed method analysis can be applied, in which the variance related to assignments and to individuals is estimated simultaneously. Hence, in statistical testing both variance components are taken into account (therefore type I errors will be minimized).

Writing processes as dependent variables. In some studies, changes in writing processes due to experimental pedagogies are studied explicitly. It has been shown that not only do experimental pedagogies influence the orchestration of writing processes and that these differences in orchestration are related to text quality (Braaksma et al., 2004). Showing which processes are affected by instruction, and how these processes are related to the quality of the resulting texts, opens the so-called 'black box'. Such studies show not only effects of interventions on (aspects of) text quality, but also effects on the underlying processes. Hence, these studies contribute to the understanding of the effects of such interventions as well as to theory (see also, Garcia & Fidalgo, 2008).

Learning process characteristics. Few studies include implementation measures even though this might be very informative. A simple inclusion of time on task measures might show that those students involved more often in dialogues progress more than students who were relatively less often involved in this kind of learning activities (Frijters, Ten Dam & Rijlaarsdam, 2008). Researchers may also include the quality of student's work during the intervention as a factor to explain variance in results (Braaksma, Van den Bergh, Rijlaarsdam, & Couzijn, 2001).

Increasing statistical power and creating replications. Especially in cases where a control group controls for maturation effects, an ethical problem arises: participants in the control group often do not receive the most effective treatment. Adding one extra measurement solves this problem and affects the statistical power and possibilities for interpretation considerably. An instance is a so-called swapping panel design (Shadish, Cook & Campbell, 2001). In such a design both groups are tested three times: pretest, midtest and posttest. However, the first group takes the experimental manipulation between the pre- and midtest, whereas the second group takes the experimental manipulation between mid- and posttest. As such, two experiments are in fact combined into one.

Concluding remarks

All in all there are many different ways to strengthen writing research. This holds for methodological aspects as well as for theoretical aspects of writing research. Please note that many of the suggested methodological improvements not only serve the generalizability (or internal validity) of studies, but also increase possibilities for interpretations of research findings. New methods of study, (the help of) new techniques, will contribute to better measure text quality and writing processes, but only if theoretical advancements are made. Data do not speak. Data are silent. They tell us nothing without a theoretical interpretation.

Writing research agenda

One conclusion is that many of the questions posed in this introduction need attention from researchers. Although much progress has been made, our impression is that we are not well prepared for the development of future writing curricula. We must redefine 'writing' as something along the lines of 'document design' or' information design' allowing multimodal solutions for information exchange. Research effort must be invested in supporting the development of productive information design capacities in children en adolescents, in a context of intercultural and multilingual communication, with all kinds of technology.

Technology will support the formulation process in any language we want to communicate in writing (word prediction, clause prediction); including other modalities of communication in digital documents will be easier. However, while technology supports the writer, it also may increase the cognitive burden of writing and learning-to-write. Writing has always been called a complex activity, and that is what it will continue to be: designing documents for communication and learning will become more demanding, cognitively speaking, in a world where the definition of what a good text entails becomes increasingly complex and vague, as a result of variation in input and as a result of globalization of information, interculturalization and multilingualism.

Review studies of invention studies offer us an increasing insight into/knowledge about what works in writing education. What we need in this domain is a more fine-grained analysis of

interventions, and more elaborated theoretical accounts of choices in the design of the intervention. In the domain of writing processes, there is increasing knowledge about writing processes, their constituent activities, and the interaction between them. Additionally we also know more about which skills contribute to the quality of the process and about how adaptivity to contexts (within and between writing tasks) distinguishes the better writer from the weaker writer. New research technology will make it possible to study larger groups of participants, allowing researchers to include more relevant variables in the analysis such as writer characteristics, subskill performances, writing processes and text characteristics. It will also enable researcher to collect more tasks per participant which is extremely important to distinguish variances due to task and to subject. In the domain of writing-to-learn, a promising start has been made to build a theoretical account about why and under what conditions the act of writing contributes to learning. Here we need a robust theory and empirical studies that relate types of learning tasks to types of writing acts.

However, a neglected domain of research remains the underlying mechanisms of learning or skill acquisition. In intervention studies, the focus is on the effect of arrangements of learning activities on outcomes, in terms of performance and sometimes on processes underlying this performance. What the students actually do, which cognitive operations are involved when students are involved in learning-to-write activities, however, remains unknown. How writing skill develops, in various ways, in various learners, under various conditions, is still a black box. Designing theories of learning and empirical studies to test these theories in writing-to-learn and learning-to-write situations, might contribute to relate the domains of knowledge we presented in this chapter.

References

Ackerman, J.M. (1993). The promise of writing to learn. *Written Communication*, *10*, 334-370.

Alamargot, D., & Chanquoy, L. (2001). *Studies in Writing, vol. 9. Through the models of writing*. Dordrecht: Kluwer Academic Publishers.

Alamargot, D., & Fayol, F. (2009). Modelling the development of written composition. In R. Beard,

D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (p. 23-47). Los Angeles: SAGE.

Anderson, J.R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261-295.

Applebee, A.N. (1984). Writing and reasoning. *Review of Educational Research*, *54*, 577-596.

Applebee, A.N., & Langer, J.A. (2006). *The stage of writing instruction in American schools: What existing data tell us.* Report. Center on English Learning & Achievement. Albany, NY: State University of New York.

Bacha, N. (2001). Writing evaluation: What can analytic versus holistic essay scoring tell us? *System*, *29*, 371–383.

Baddeley, A.D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General, 129*, 126-145.

Baker S., Gersten, , R., & Graham, S. (2003). Teaching expressive writing to students with learning disabilities: Research-based applications. *Journal of Learning Disabilities*, 36, 109-123.

Bangert-Drowns, R.L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research, 63*, 69-93

Bangert-Drowns, R.L., Hurley, M.M., & Wilkinson, B. (2004). The effects of school-based writing to learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58.

Barkaoui, K. (2007). Rating scale impact on EFL esay marking: A mixed-method study. *Assessing Writing*, *12*, 86-107.

Baron, N. (2006). The future of written culture: Envisioning language in the New Millennium, *Iberuca, 9*, 7-31

Bazerman, C. (Ed.) (2007). *Handbook of research on writing. History, society, school, individual, text*. London: Routledge.

Beard, R., Myhill, D., Riley. J., & Nystrand, M. (Eds.) (2009). *The SAGE handbook of writing*

development. Los Angeles: SAGE.

Belanoff, P., & Elbow, P. (1986). Using portfolios to increase collaboration and community in a writing program. *Writing program administration, 9*(3), 27-40.

Bereiter C., & M. Scardamalia (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berner, A., Boswell, W., & Kahan, N. (1996). Using the tape recorder to respond to student writers. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Effective teaching and learning of writing* (pp. 339-357). Amsterdam: Amsterdam University Press.

Berninger, V.W., Abbott, R.D., Augsburger, A., & Garcia, N. (2009). Comparison of pen and keyboard transcription modes in children with and without learning disabilities. *Learning Disability Quarterly*, *32*, 123-141.

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*, 166-195.

Biehal, G. J. , & Chakravarti, D. (1989).The effects of concurrent verbalization on choice processing. *Journal of Marketing Research*, *26*, 84-96.

Black, R.W. (2009). Online fan fiction, global identities, and imagination. *Research in the Teaching of English, 43*, 397-425

Boscolo, P., & Carotti, L. (2003). Does writing contribute to improving high school students' approach to literature? *L1 – Educational Studies in Language and Literature*, *3*(3), 197-224.

Boscolo, P. (2009). Engaging and motivating children to write. In: R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 300-312). Los Angeles: SAGE.

Boscolo, P., & Mason, L. (2001). Writing to learn, writing to transfer. In P. Tynjälä, L. Mason, & K. Lonka (Volume Eds.). *Studies in Writing, Volume 7, Writing as a learning tool: Integrating theory and practice* (pp. 83-104). Dordrecht: Kluwer Academic Press.

Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted? *Educational Technology &*

Society, *12*, 257-268.

Braaksma, M.A.H., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94,* 405-415.

Braaksma, M. A. H., Van den Bergh, H., Rijlaarsdam, G., & Couzijn, M. (2001). Effective learning activities in observation tasks when learning to write and read argumentative texts. *European Journal of Psychology of Education, 16*, 33-48.

Braaksma, M. Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B.H.A.M. (2004). Observational learning and its effects on the orchestration of the writing process. *Cognition & Instruction*, *22*, 1-36.

Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction*, *12*, 103-123.

Breland, H.M., & Jones, R.J. (1984). Perceptions of writing skills. *Written Communication*, *1,* 101-119.

Brennan, R.L. (1992). *Elements of generalizability theory*. Iowa City, IA: ACT.

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *School councils research studies: The development of writing abilities (11-18).* London: Macmillan Education.

Butterworth B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research, 4*, 75 - 8

Chambliss, M.J., Christenson, L.A., & Parker, C. (2003). Fourth graders composing scientific explanations about the effects of pollutants: Writing to understand. *Written Communication*, *20*, 426-454.

Chapman, M. (2006). Research in writing: Pre-school through elementary, 1984-2003. *L1 -- Educational Studies in Languages and Literature*, *6*(2), 7-27.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English*, *18*, 65-81.

Chatel, R.G. (2002). New technology, new literacy: Creating a bridge for English language learners.

The New England Reading Association Journal, 38(3), 45-49.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*, 409-426.

Cizek, G.J., & Page, B.A. (2003). The concept of reliability in the context of automated scoring. In M.D. Shermis, & J.C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 125-145). Mahwah, NJ: Lawrence-Erlbaum.

Clark, H. (1973). The language-as-a-fixed-effect Fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

Coffman, W.E. (1966). On the validity of essay tests of achievement. *Journal of Educational Measurement, 3*, 151 - 156.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum

Cooke, L. & Cuddihy, E. (2005, July). *Using eye-tracking to address limitations in think-aloud protocol*. Professional Communivation Conference (july 10-13), IPCC Proceedings international.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

Crystal, D. (2003). *English as a Global Language*. Cambridge: Cambridge University Press.

Culpepper, M.M., & Ramsdale, R. (1982). A comparison of multiple choice and essay test of writing skills. *Research in the Teaching of English*, *16*, 295-297.

Cushing Weigle, S. (1994). Effects of training on raters of ESL compositions. *Language Testing*, *11*, 197-223.

Cushing Weigle, S. (1998). Using FACETS to model rater training effects. *Language Testing, 15*, 263-287.

Cushing Weigle, S. (2002). *Assessing writing.* Cambridge: Cambridge University Press.

Daiute, C. (1986). Physical and cognitive factors in revising: Insights from studies with computers. *Research in the Teaching of English*, *20*, 141-159.

Davidson-Shivers, G.V., Nowlin, B., & Lanouette, M. (2002). Do multimedia lesson structure and learning styles influence undergraduate writing performance? *College Student Journal*, *36*(1), 20-32.

Davis, A., & McGrail, E. (2009). "Proof-Revising" with podcasting: Keeping readers in mind as students listen to and rethink their writing. *Reading Teacher*, *62*(6), 522-529.

De Groot, A.D. (1946). *Het denken van den schaker. Een experimenteel-psychologische studie. [The thinking of the chess player. An experimental psychological study]*. Noord-Hollandsche Uitgevers Maatschappij, Amsterdam.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*, 391-407.

Deremer, M.L. (1998). Writing assessment: Raters' elaboration of the rating task, *Assessing Writing*, *5*, 7-29.

Diederich, P.B., French, J.W., & Carlton, S.T. (1961). *Factors in the judgment of writing ability.* Princeton, NJ: Educational Testing Service.

East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, *14*, 88-115.

Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, *26*, 156-185.

Elder, C., Knoch, U., Barkhuizen, G., & von Randow, J. (2005). Individual feedback to enhance rater training: Does it work? *Language Assessment Quarterly*, *2*, 175-196.

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, *28*, 122-128.

Ericsson K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data.* MIT Press: Cambridge, MA.

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology, 92,* 694-702.

Ferretti, R., Lewis, W., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology*, *101*, 577-589.

Fitzgerald, J. (2006). Multilingual writing in preschool through 12th Grade: The last 15 years. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of writing research (pp. 337-354). New York/London: The Guilford Press.

Flower, L., & Hayes, J.R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32,* 365-87.

Frijters, S., Ten Dam, G., & Rijlaarsdam, G. (2008). Effects of dialogic learning on value-loaded critical thinking. *Learning and Instruction*, *18*, 66-82

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science*, *21*, 45-72.

Galbraith, D. (1996). Self-monitoring, discovery through writing, and individual differences in drafting strategy. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 121- 141). Amsterdam: University Press.

Galbraith, D. (1999). Writing as a knowledge-constituting process. In D. Galbraith, & M. Torrance (Vol. Eds.), *Studies in writing: Vol. 4. Knowing what to write. Conceptual processes in text production* (pp.139-159). Amsterdam: Amsterdam University Press.

Galbraith, D., & G. Rijlaarsdam (1999). Effective strategies for the teaching and learning of writing. *Learning and Instruction*, *9*, 93-108.

Galbraith, D. (2009). Writing about what we know: Generating ideas in writing. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 48-64). Los Angeles etc.: SAGE.

Garcia, J., & Fidalgo, R. (2008). Orchestration of writing processes and writing products: A comparison of sixth-grade students with and without learning disabilities. *Learning Disabilities: A Contemporary Journal, 6*, 77-98.

Gebril, A. (2009). Score generalizability of academic writing tasks: Does one method fit it all? *Language Testing*, *26*, 507-531.

Gersten, R., & Baker, S. (2001). Teaching expressive writing to students with learning disabilities: A meta-analysis. *The Elementary School Journal*, *101*, 251-272.

Glaser, C., Kessler, C., & Brunstein, J.C. (2009). Improving fourth graders' self-regulated writing skills: Effects on strategy-related, holistic, and subjective measures of writing competence. *Zeitschrift für Pädagogische Psychologie, 23*, 5-18.

Graham, S. (2006). Strategy instruction and the teaching of writing. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187-207). New York/London: The Guilford Press.

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist*, *35*, 3-12.

Graham, S., & Hebert, M. (2010). *Writing to read; Evidence for how writing can improve reading*. *A report from Carnegie Corporation of New York*. Washington DC: Alliance for Excellent Education.

Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*, 445-476.

Graham, S., & Perin, D. (2007b). What we know, what we still need to know: Teaching adolescents to write. *Scientific Studies of Reading*, *11*, 313-335.

Grejda, G.F., & Hannafin, M.J. (1992). Effects of word-processing on 6th graders holistic writing and revisions. *Journal of Educational Research*, *85*, 144-149

Groenendijk, T,. Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2008). How do secondary school students write poetry? How creative writing processes relate to final products. *L1 –*

Educational Studies in Language and Literature, *8*(3), 57-80.

Gunel, M., Hand, B., & Prain,V. (2007). Writing for learning in science: A secondary analysis of six studies. *International Journal of Science and Mathematics Education*, *5*, 615-637.

Hamp-Lyons, L. (1995). Research on the rating process. Rating nonnative writing: the trouble with holistic scoring. *TESOL Quarterly, 29*, 759-762.

Hamp-Lyons, L. (2007). Worrying about rating. *Assessing Writing*, *12*, 1-9.

Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice theory and research*. Cresskill, NJ: Hampton Press.

Hand, B., Gunel, M., & Ulu, C. (2009). Sequencing embedded multimodal representations in a writing to learn approach to the teaching of electricity. *Journal of Research in Science Teaching, 46*, 225-247.

Hand, B., Wallace, C., & Yang, E. (2004). Using the science writing heuristic to enhance learning outcomes from laboratory activities in seventh grade science: Quantitative and qualitative aspects. *International Journal of Science Education*, *26*, 131-149.

Harris, K. R., Graham, S., Mason, L., & Friedlander, B. (2007). *Powerful writing strategies for all students*. Baltimore, MD: Brookes.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing. Theories, methods, individual differences and applications* (pp. 1- 27). Mahwah, NJ: Lawrence Erlbaum Associates.

Hayes, J. R., & Flower, L.S. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum

Hayes, J.R., & Flower, L.S. (1986). Writing research and the writer. *American psychologist*. *41*, 1106-1113.

Hillocks, G. (1986). *Research on written composition: New directions for teaching*. Urbana, Il: ERIC Clearinghouse on Reading and Communicative Skills.

Hillocks, G. (2006). Research in writing, Secondary School, 1984-2003. *L1 – Educational Studies in Languages and Literature*, 6(2), 29-52.

Holliway, D.R., & D. McCutchen (2004). Audience perspective in young writers' composing and revising. Reading as the reader. In L. Allal, L. Chanquoy, & P. Largy (Vol.Eds.), *Studies in Writing. Vol. 13: Revision: Cognitive and Instructional Processes* (p. 87-101). Boston etc.: Kluwer Academic Publishers.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, *60*, 237-263.

Jacoby, L.L., Lindsay, D.S., & Toth, J.P. (1992). Unconsious influences revealed: Attention, awareness and control. *American Psychologist*, *47*, 802-808.

Janssen, D., Van Waes, L., Van den Bergh, H. (1996). Effects of thinking akoud on writing processes. In C.M. Levy, & S.E. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 233-250). Mahwah, NJ: Lawrence Erlbaum Associates.

Jensen, G.H., & DiTiberio, J.K. (1984). Personality and individual writing processes. *College Composition and Communication*, *35*, 285-300.

Jorge-Botana, G., Leon, J.A., Olmos, R., & Escudero, I. (2010). Latent semantic analysis parameters for essay evaluation using small-scale corpora. *Journal of Quantitative Linguistics*, *17*, 1-29.

Joseph, L.M., & Konrad, M. (2009). Teaching students with intellectual or developmental disabilities to write: A review of the literature. *Research in developmental disabilities, 30,* 1-19.

Kellogg, R. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning Memory and Cognition, 14*, 355-365.

Kellogg, R.T. (1999). *The Psychology of writing*. NewYork: Oxford University Press.

Kellogg, R. (2001). Competition for working memory among writing processes. *American Journal of Psychology, 114*, 175-191.

Kellogg, R.T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of*

Writing Research, *1*(1), 1-26.

Keys, C.W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, *36*, 1065-1084.

Kieft, M., Rijlaarsdam, G., Galbraith, D., & Van den Bergh, H. (2007). The effects of adapting a writing course to students' writing strategies. *British Journal of Educational Psychology*, *77*, 565-578.

Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2008). An aptitude–treatment interaction approach to writing-to-learn. *Learning and Instruction*, *18*, 379-390.

Klein, P.D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, *11*, 203-270.

Klein, P.D. (2000). Elementary students' strategies for writing-to-learn in science. *Cognition and Instruction*, *18*, 317-348.

Klein, P.D. (2004). Constructing scientific explanations through writing. *Instructional Science*, *32*, 191-231.

Klein, P. D., Boman, J. S., & Prince, M. P. (2007). Developmental trends in a writing to learn task. In M. Torrance, D. Galbraith, & L. van Waes (Vol. Eds.), *Studies in Writing*, *Vol. 20*, *Writing and Cognition: Research and Application* (pp. 201-217). Amsterdam: Elsevier.

Klein, P. D., Piacente-Cimini, S., & Williams, L. A. (2007). The role of writing in learning from analogies. *Learning and Instruction, 17*, 595-611.

Klein, P.D., & Kirkpatrick, L.C. (2010). A framework for content area writing: Mediators and moderators. *Journal of Writing Research, 2*(1), 1-46.

Knoch, U., Read, J., & von Randow, J. (2007). Re-training writers online: How does it compare with face-to-face training? *Assessing Writing*, *12*, 26-43.

Kolberg, P., & Severinson Eklundh, K. (2003). Studying writers' revising patterns with S-notation. In T. Olive, & C.M. Levy (Vol. Eds.), *Studies in Writing, vol. 10. Contempory tools and*

techniques for studying writing (p. 89-104). Dordrecht, NL.: Kluwer Academic Publishers.

Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese L2 writing performance. *Language Testing*, *19*, 3-31.

Kucan, L., & Beck, I.L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, *67*, 271-299

Kuhlemeier, H., & Van den Bergh, H. (1998). Relationships between language skills and task effects. *Perceptual and Motor Skills*, *86*, 443-486.

Landauer, T., Laham, D., & Foltz, P. (2000). The intelligent Essay Assessor. *IEEE Intelligent Systems*, *15*(5), 27-31.

Langer, J.A., & Applebee, A.N. (1987). *How writing shapes thinking: A study of teaching and learning.* Urbana, Ill.: National Council of Teachers of English.

Lehmann, R.H. (1990). Reliability and generalizability of ratings of compositions. *Studies in Educational Evaluation, 16*, 501-512.

Leow, R.P., & Morgan-Short, K. (2006). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, *26*, 35-57.

Levy, C. M., Marek, J. P, & Lea, J. (1995). Concurrent and retrospective protocols in writing research. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.). *Writing research: Theories, models and methodology (pp.* 542-556). Amsterdam: Amsterdam University Press.

Lindgren, E. (2004). The uptake of peer-based interventions in the writing classroom. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Vol. Eds.), *Studies in Writing*. Vol 14. *Effective learning and teaching of writing*. 2nd Edition, Part 2: Studies in how to teach writing. (pp. 259-274). New York: Kluwer Academic Publishers

Lloyd-Jones, R. (1977). Primary-trait scoring. In C.R. Cooper, & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 33-68). Urbana, IL: National Council of Teachers of English.

Lumbelli, L., & Paoletti, G. (2004). Monitoring local coherence through bridging integration. In G.

Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Studies in writing. Vol. 14, Effective learning and teaching of writing*. (2nd ed. p. 197-208). Boston: Kluwer Academic Publishers.

Lumley, T. (2002) Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, *19*, 246-274.

MacArthur, C.A., Graham, S., & Fitzgerald, D. (Eds.).(2006). *Handbook of writing research*. New York/London: The Guilford Press.

MacArthur, C. (2005). Assistive technology for writing tools for struggling writers. In L. van Waes, M. Leijten, & C.Neuwirth (Vol. Eds.), *Studies in Writing: Vol 17.Writing and digital media* (pp. 11-20).Oxford: Elsevier.

MacArthur, C.A. (2006). The effects of new technologies on writing and writing processes. In C.A. MacArthur, S. Graham, & D. Fitzgerald (Eds.), *Handbook of writing research* (pp. 248-262). New York/London: The Guilford Press.

MacArthur, C.A., & Cavalier, A.R. (2004). Dictation and speech recognition technology as test accommodations. *Exceptional Children, 7*, pp. 43-58

Marshall, J. (1987). The effects of writing on students' understanding of literary texts. *Research in the Teaching of English, 21,* 30–63.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*, 299-325.

McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 115-130). New York/London: The Guilford Press.

Meuffels, B. (1994). *De verguisde beoordelaar. Opstellen over opstelbeoordeling*. [The abused assessor. Essays about the evaluation of essays]. Thesis Publishers: Amsterdam.

Moss, P. (1994). Can there be validity without reliability? *Educational Researcher*, *23*, 5-12.

Murphy, J.J. (Ed.).(2001). *A short history of writing instruction: From ancient Greece to modern America*. (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Myford, C.M., & Wolfe, E.W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, *4*, 386-422.

Myford, C.M., & Wolfe, E.W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, *5*, 189-227.

Neuwirth, C. M., Kaufer, D. S., Chandhok, R., & Morris, J. H. (1990). Issues in the design of computer-support for co-authoring and commenting. In *Proceedings of the Third Conference on Computer-Supported Cooperative Work* (CSCW '90) (pp. 183-195). Baltimore, MD: Association for Computing Machinery.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall

Newell, G.E. (2006). Writing to learn: How alternative theories of school writing account for student performance. In C.A. MacArthur, S., Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 235-247). New York / London: The Guilford Press.

Newell, G.E., & Winograd, P. (1995). Writing about and learning from history texts: The effects of task and academic ability. *Research in the Teaching of English*, *29*, 133-163.

Nixon, H. (2003). New research literacies for contemporary research into literacy and new media? *Reading Research Quarterly, 38*, 407-413

Nuvoli, G. (2000). Revising of texts with word-processing. *Psychological Reports*, *87*, 1139-1146.

Nystrand, M., Gamoran, A, & Carbonaro, W. (2001). On the ecology of classroom instruction: The case of writing in high School English and Social Studies. In P. Tynjala, L. Mason, & K. Lonka (Vol. Eds.), *Studies in Writing. Vol 7. Writing as a learning tool; Integrating theory and practice* (pp. 57-81). Boston: Kluwer Academic Publishers

Ochsner, R., & Fowler, J. (2004). Playing devil's advocate: Evaluating the literature of the WAC/WID movement. *Review of Educational Research, 74*, 117-140.

OECD (2009). *Education at a glance 2009. OECD indicators.* Indicators D1.2a and D1.2b. Retrieved at www.oecd.org on Feb. 2nd, 2010.

Olive, T., Kellogg, R.T., & Piolat, A. (2001). The triple task technique for studying the processes of

writing: Why and how? In T. Olive, & C. M. Levy (Vol. Eds.), *Studies in Writing, vol. 10. Contemprorary tools and techniques for studying writing* (31-59). Dordrecht: Kluwer Academic Publishers.

Oliver, R. (2004). Writing; in your own words: Children's use of information sources in research projects. In G. Rijlaarsdam, H. Van den Bergh, & M. Couzijn (vol. Eds.), *Studies in Writing. Vol 14*. *Effective learning and teaching of writing*. 2nd Edition, Part 2: Studies in how to teach writing (pp. 367-380). New York: Kluwer Academic Publishers

Pérez-Sabater, C., Peña-Martínez, G., Turney, E., & Montero-Fleta, B. (2008) A spoken genre gets written: Online football commentaries in English, French and Spanish. *Written Communication, 25*, 235-261

Piolat, A., Kellogg, R.T., & Farioli, F., (2001). The triple task technique for studying writing processes on which task attention is focussed? *Current Psychology Letters*, *4*, 67-83.

Piolat, A., Roussey, J.Y, Olive, T., & Farioli, F. (1996). Charge mentale et mobilisation des processus rédactionnels: Examen de la procédure de Kellogg. *Psychologie Française*, *41*, 339-354.

Pugalee, D.K. (2004) A comparison of verbal and written descriptions of students' problem solving processes. *Educational Studies in Mathematics*, *55*, 27-47

Purcell-Gates, V., Duke, N. K., & Martineau, J. A. (2007). Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly, 42*, 8-45.

Quellmalz, E.S. (1982). Designing writing assessments: Balancing fairness, utility and cost. (CSE Report]. University of California, LA.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413-425.

Raedts, M., Rijlaarsdam, G., Van Waes, L., & Daems, F. (2007). Observational learning through video-based models: Impact on students' accuracy of self-efficacy beliefs, task knowledge and writing performances. In P. Boscolo, & S. Hidi (Vol Eds.), *Studies in Writing, Vol. 19, Writing*

and Motivation (pp 219-238). Oxford: Elsevier

Ransdell, S. (1993). Generating thinking-aloud protocols: Impact on the narrative writing of college students. *The American Journal of Psychology*, *108*, 89-98.

Ransdell, S., & Levy, C. M. (1996). Working memory constraints on writing quality and fluency. In C. M. Levy, & S. E. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 93-106). Mahwah, NJ: Lawrence Erlbaum Associates.

Ransdell, S., Levy, C. M., & Kellogg, R.T. (2002). The structure of writing processes as revealed by secondary task demands. *L1 -- Educational Studies in Language and Literature, 2*(2), 141-163.

Rau, P. S., & Sebrechts, M. M. (1996). How initial plans mediate the expansion and resolution of options in writing. *The quarterly Journal of Experimental Psychology*, *49*, 616-638.

Rijlaarsdam, G., & Couzijn, M. (2000). Writing and learning to write: A double challenge. In: R.J. Simons (Ed.), *New learning* (pp.157-189). Dordrecht, NL: Kluwer.

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Raedts, M., Van Steendam, E., Toorenaar, J., & Van den Bergh, H. (2008). Observation of peers in learning to write: Practice and research. *Journal of Writing Research, 1*(1*),* 53-83.

Rijlaarsdam, G., Couzijn, M., Janssen, T., Braaksma, M., & Kieft, M. (2006). Writing experiment manuals in science education: The impact of writing, genre, and audience. *International Journal of Science Education, 28*, 203–233.

Rijlaarsdam, G., Van den Bergh, H., & Couzijn, M. (Eds.) (2004). *Studies in Writing: Vol. 14. Effective Learning and Teaching of Writing. A Handbook of Writing in Education*. Second edition. Boston: Kluwer Academic Publishers.

Rivard, L.P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching*, *31*, 969-983.

Rivard, L.P. (2004). Are language-based activities in science effective for all students, including low achievers? *Science Education, 88*, 420-442.

Roca de Larios, J., Murphy, L., & Marin, J. (2002). A critical examination of L2-writing process

research. In S. Ransdell, & M.L. Barbier (vol. Eds.), *Studies in Writing*, *Vol. 11*: *New directions for research in L2 writing* (pp. 11-47]. Dordrecht, NL: Kluwer Academic Publishers.

Rogers, L.A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology, 100,* 879-906.

Russell, D.R., Lea, M., Parker, J., Street, B., & Donahue, T. (2009). Exploring notions of genre in "Academic literacies" and "Writing across the curriculum": Approaches across countries and contexts. In C. Bazerman, A. Bonini, & D. Figueiredo (Eds.), *Genre in a changing world* (pp. 395-424). Fort Collins, Colorado: The WAC Clearinghouse. West Lafayette, Ind.: Parlor Pres.

Russo, J.E., Johnson, E.J., & Stephans, D.L. (1989) The validiy of verbal protocols. *Memory & Cognition*, *17*, 759-769

Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology, 97*, 43-54.

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. Wittrock (Ed.), *Handbook of research on teaching* [3rd ed.] (pp. 778-803). New York: McMillan.

Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modelling. *Language Testing*, *22*, 1-30.

Schoonen, R., Vergeer, M., & Eiting, M. (1997). The assessment of writing ability: Expert readers versus lay readers. *Language Testing*, *14*, 157-184.

Schriver, K.A. (1997). *Dynamics in document design*. New York: Wiley.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2001). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Shermis, M.D., Burnstein, J., & Leacock, C. (2006). Applications of computers in assessment and analysis of writing. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*. (403-416). New York/London: The Guilford Press.

Schilperoord, J. (1996). It's about time: Temporal aspects of cognitive processes in text processing. Amsterdam: Rodopi.

Siemer, M., & Joorman, J. (2003). Power and measures of effect size in analysis of variance with fixed and random nested factors. *Psychological Methods, 8*, 497-517.

Smagorinsky, P. (2006). Writing in school contexts: Introduction. *L1 -- Educational Studies in Language and Literature*, *6*(2), 1-7.

Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Stratham, J.F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud pro-tocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (p. 89–112). London: Sage.

Strömqvist, S., Holmqvist, K., Johansson, V., Karlsson, H., & Wengelin, Å. (2006). What key-logging can reveal about writing. In K. Sullivan, & E. Lindgren (Eds.), *Studies in Writing, vol. 18. Computer key-stroke logging and writing: methods and applications* (p. 45-72). Boston: Kluwer Academic Publishers

Thaiss, C., & Porter, T. (2010). The State of WAC/WID in 2010: Methods and results of the U.S. surveys of the international WAC/WID mapping project. *College Composition and Communication, 61*, 534-570.

Tirkkonen-Condit, S. (2005). Process research: State of the art and where to go next. *Across Languages and Cultures, 3*, 5-19.

Torrance, M., Fidalgo, R., & García, J.N. (2007). The teachability and effectiveness of cognitive self-regulation in sixth-grade writers. *Learning and Instruction, 17*, 265-285.

Torrance, M., Thomas, G.V., & Robinson, E.J. (2000). Individual differences in undergraduate essay writing strategies: A longitudinal study. *Higher Education*, *39*, 181-200.

Troia, G.A. (2008). Writing instruction for students with learning disabilities. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (p. 324-336). New York:

Guilford Press.

Tynjälä, P. (2001). Writing, learning and the development of expertise in higher education. In P. Tynjälä, L. Mason, & K. Lonka (Vol. Eds.). *Studies in Writing, Volume 7, Writing as a learning tool: Integrating theory and practice* (pp. 37-56). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Uppstad, P.H., & Solheim, O.J. (2006). Aspects of fluency in writing. *Journal of Psycholinguistic Research*, *36*, 79-87.

Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing and translation from L1 into L2. *Journal of Second Language Writing*, *5*, 271–294.

Van den Bergh, H., & Rijlaarsdam, G. (2001). Changes in cognitive activities during the writing process, and relations with text quality. *Educational Psychology*, *21*, 373-385.

Van den Bergh, H. (1988). Schrijven en schrijven is twee: Een onderzoek naar de samenhang tussen prestaties op verschillende schrijftaken. [Writing and writing makes two: An investigation into the relation between performances on different writing tasks.] *Tijdschrift voor Onderwijsresearch, 13*(6), 311-324.

Van den Bergh, H. (1989). Functionele schrijfopdrachten en stelopdrachten: Een onderzoek naar de construct-validiteit van schrijfvaardigheidsmetingen. [Functional writing assignments and composition assignments: An investigation into the construct validity of writing proficiency measures.] *Tijdschrift voor Onderwijsresearch, 14*, 151-171.

Van den Bergh, H., & Rijlaarsdam, G. (2007). The dynamics of idea generating during writing: An online study. In M. Torrance, L. van Waes, & D. Galbraith (Vol. Eds.), *Studies in Writing Vol. 20, Writing and Cognition: Research and Applications* (pp. 125 - 150). Amsterdam: Elsevier.

Van den Bergh, H., Rijlaarsdam, G., & Zwarts, M. (1992). Incidentele transfer: Een aanzet tot een baseline. [Incidental transfer: A baseline study.] *Tijdschrift voor Onderwijsresearch*, *17*, 55-66.

Van den Bergh, H., Rijlaarsdam, G., Janssen, T., Braaksma, M., van Weijen., D., & Tillema, M.

(2009). Process execution of writing and reading: Considering text quality, learn and task characteristics. In M.C. Shelley II, L.D. Yore, & B. Hand (eds.), *Quality research in literary and science education* (pp 399-425). Boston; Springer.

Van den Bergh, H., & Eiting, M. (1989). A method of estimating rater reliability. *Journal of Educational Measurement*, *26*, 29 – 40.

Van der Hoeven, J. (1997). *Children's composing. A study into the relationships between writing processes, text quality, and cognitive and linguistic skills*. Utrecht Studies in Language and Communication, Vol. 12. Amsterdam: Rodopi.

Van Steendam, E., Rijlaarsdam, G., Sercu. L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning & Instruction*, *20*, p 316-327.

Van Waes, L. & Leijten (2006). Logging writing processes with inputlog. In, L. Van Waes, M. Leijten, & C.M. Neuwirth (Eds.), *Studies in Writing, Vol 17: Writing and digital media* (pp. 150-165). Amsterdam: Elsevier.

Van Waes, L, Leijten, M., & Van Weijen, D. (2009, in press). Keystroke logging in writing research. Observing writing processes with Inputlog. *GFL-German as a foreign language* - ISSN 1470-9570.

Van Weijen, D. (2009). *Writing processes, text quality, and task effects: Empirical studies in first and second language writing*. Utrecht, the Netherlands: LOT Dissertation Series.

Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2008). Differences in process and process-product relations in L2 writing. *International Journal of Applied Linguistics*, *156*, 203-226.

Van Weijen D., Van den Bergh H., Rijlaarsdam G., & Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. *Journal of Second Language writing*, *18*(4), 235-250.

Vaughan, C. (1991). Holistic assessment: What goes on in the rater's mind? In L. Hamp-Lyons (Ed.),

Assessing second-language writing in academic contexts (pp. 111-125). Norwood, NJ: Ablex.

Verheyden, L. (2010). *Development in writing in grades 3 and 4. Four studies on factors that contribute to explain growth patterns*. Dissertation University of Leuven, Belgium.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research*, *10*, 157-180.

Wengelin, A., Leijten, M., & Van Waes, L. (2010). Studying reading during writing: New perspectives in research. *Reading and Writing: An Interdisciplinary Journal*, 23, pp735-742..

Whitaker D., Berninger, V., Johnston, J, & Swanson. H.L. (1994). Intraindividual differences in levels of language in intermediate grade writers – implications for the translating process. *Learning and Individual Differences, 6*, 107-130.

Wolfe, E.W. (1997). The relationship between essay reading style and scoring proficiency in a psychometric scoring system. *Assessing Writing*, *4*, 83-106.

Woodall, B. R. (2002). Language-switching: Using the first language while writing in a second. *Journal of Second Language Writing*, *11*, 7–28.

Yore, L.D., Bisanz, G.L., & Hand, B.M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education, 25*, 689-725.

Zimmerman, B., & A. Kitsantas (2007). A writer's discipline: The development of self-regulatory skill. In P. Boscolo, & S. Hidi (Vol. Eds.), *Studies in Writing: Vol. 19, Writing and Motivation* (p. 51-69). Oxford: Elsevier.

Table 1. Three macro-stages of writing development (Adapted from Kellogg, 2008)

	Knowledge telling	*Knowledge transforming*	*Knowledge crafting*
Aim of writing:	Conveying 'what one knows' in written form	Transforming 'what one knows' for the author's benefit	Crafting 'what one knows' for the reader's benefit
Text results from interaction among:	Author's ideas	Author's ideas and written text	Author's ideas, written text and presumed readers' representations
Scope of planning:	Idea retrieval, formulating propositions	Same; plus an ongoing selection and evaluation of ideas and text before committing them to paper	Same; but selection and evaluation occur in the light of how readers will interpret and value the text
Scope of reviewing:	(At best) adequate translation of ideas into well-formed sentences	An interaction between reviewing the text and reviewing one's ideas and intentions	Same; expanded by presumed readers' representations of the text and/or the ideas

Table 2. Summary of Experimental-Control effect size statistics for writing intervention treatments

Treatment	Number of studies	Grade range	Weighed mean effect
Process approach	21	4-12	.32
Explicit teaching			
Grammar	11	4-11	-.32
Sentence Combining	5	4-11	.50
Strategy Instruction	20	4-10	.82
Summarization	4	4-12	.80
Scaffolding students' writing			
Prewriting	5	4-9	.32
Inquiry	5	7-12	.32
Collaborative writing	7	4-12	.75
Study of models	6	4-12	.25
Setting product goals	5	4-8	.70
Alternative mode: Word processing	18	4-12	.55