Journal of Educational Psychology
2002, Vol. 94, No. 2, 405–415

0022-0663/02/$5.00 DOI: 10.1037//0022-0663.94.2.405

Observational Learning and the Effects of Model–Observer Similarity

Martine A. H. Braaksma
University of Amsterdam

Gert Rijlaarsdam and Huub van den Bergh
University of Amsterdam and Utrecht University

This study examined the effects of similarity in competence between model and observer on the effectiveness of observational learning in argumentative writing. Participants ($N = 214$, 8th grade, mixed ability) were assigned to 1 of 3 conditions: an observation/weak-focus, an observation/good-focus, or a control condition. The two observational-learning groups observed pairs of peer models performing writing tasks. Participants focused respectively on the noncompetent (weak) model or on the competent (good) model. The control group performed the writing tasks themselves. Results are consistent with the similarity hypothesis: Weak learners learn more from focusing their observations on weak models, whereas better learners learn more from focusing on good models.

Observing is the key learning activity in learning environments in which learners learn from models. It occurs when observers display new behaviors that prior to modeling are not demonstrated, even with motivational inducements to do so (Schunk, 1998). There is a large body of research about effects and conditions of modeling (Bandura, 1986, 1997; Rosenthal & Zimmerman, 1978; Schunk, 1987, 1991, 1995, 1998; Schunk & Zimmerman, 1997). With either teachers, adults, or students as models, observational learning has proven to be effective with students of various ages and in various school subjects, such as mathematics (Schunk & Hanson, 1985, 1989a, 1989b; Schunk, Hanson, & Cox, 1987), writing (Graham & Harris, 1994; Graham, Harris, & Troia, 1998; Schriver, 1992), speaking and listening (Sonnenschein & Whitehurst, 1983, 1984), and reading and writing (Couzijn, 1995, 1999). Observation of models also can raise observers' self-efficacy or personal beliefs about their capabilities to learn or perform behaviors at designated levels (Bandura, 1986, 1997). Schunk and his colleagues (Schunk & Hanson, 1985, 1989a; Schunk et al., 1987; see also Schunk, 1998, p. 148) reported effects of (various) models on students' self-efficacy, that in turn influenced learning and achievement.

The effectiveness of observational learning depends on a number of factors (Schunk, 1991, p. 113) including perceived similarity in competence between model and observer. Schunk (1987, 1991, 1998) reviewed studies that investigated how variations in model competence affect children's behavior. Some studies involved comparisons of mastery and coping models. Mastery models perform the task correctly. They may also verbalize statements reflecting positive attitudes and high levels of confidence and ability. Mastery models demonstrate rapid learning and make no errors. In contrast, coping models show their hesitations and errors, but gradually improve their performance and gain self-confidence. These models illustrate how determined effort and positive self-reflections may overcome difficulties.

In three studies, Schunk and his colleagues compared the observation of coping models with that of mastery models. In two of these studies (Schunk & Hanson, 1985; Schunk et al., 1987), participants were children with learning difficulties, and they were supposed to be similar to the coping models. Schunk and Hanson (1985) reported no differences between the coping and mastery conditions on any measure (perceived similarity in competence to the model, self-efficacy for learning, posttest self-efficacy, and posttest skill). Schunk and Hanson (1985, p. 320) supposed that the lack of differences between the coping and mastery conditions might be induced by the participants' prior successes with the task. In a replication study, Schunk et al. (1987) used a task with which participants had no prior success; they found that observing a coping model led to greater perceived similarity in competence to the model, higher self-efficacy for learning, and higher posttest self-efficacy and skill. In a third study, involving predominantly average performers, Schunk and Hanson (1989a) compared mastery models, coping–emotive models (i.e., coping models, as in the earlier studies), and coping-alone models (i.e., identical to coping–emotive models, but without negative statements about their own ability). Observing coping–emotive models led to the highest self-efficacy for learning, but no differences were found in participants' performances on the posttest. According to Schunk and Hanson (1989a, p. 433) this is due to the fact that—because of their prior experiences—participants judged themselves to be more competent than the observed coping–emotive models. Schunk and Hanson (1989a) assumed that by judging themselves to be more competent, coping–emotive participants overestimated their learning efficacy. This boost in self-efficacy resulting from comparison with the model causes efficacy to lose some of its predictive utility (Schunk & Hanson, 1989a, p. 433).

Other studies compared models high and low in competence. Model competence was inferred from the outcomes of modeled actions (success, failure), but also from symbols that denote com-

Martine A. H. Braaksma, Graduate School of Teaching and Learning, University of Amsterdam, Amsterdam, the Netherlands; Gert Rijlaarsdam and Huub van den Bergh, Graduate School of Teaching and Learning, University of Amsterdam, and Utrecht Institute of Linguistics, Utrecht University, Utrecht, the Netherlands.

We wish to thank Hein Broekkamp and Michel Couzijn for their thoughtful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Martine A. H. Braaksma, Graduate School of Teaching and Learning, University of Amsterdam, Wibautstraat 2-4, 1091 GM Amsterdam, the Netherlands. E-mail: braaksma@ilo.uva.nl

petence (e.g., prestige; Schunk, 1991, p. 114). In studies of observational learning of cognitive skills (see Schunk, 1987, pp. 161–162), the observation of competent models led to higher performance than did the observation of noncompetent models. In other words, models dissimilar in competence with observers exert more powerful effects on learners' behaviors.

Graham and his colleagues (Graham & Harris, 1994; Graham, Harris, MacArthur, & Schwartz, 1998; Graham, Harris, & Troia, 1998) also argued in favor of competent models. They outlined an instructional model for developing writing and self-regulation strategies among students with learning and writing problems. Part of this self-regulated strategy development (SRSD) model is a teacher who models good writing strategies (i.e., a competent model). SRSD improved students' writing performance.

Similarity in competence may be important in contexts in which children cannot readily discern the functional behavior (i.e., when they lack task familiarity, when there is no objective standard of performance, or when modeled actions are followed by neutral consequences; Schunk, 1987, p. 161). Furthermore, in studies in which model competence served as a source of information for purposes of self-evaluation, children were influenced by models of similar (same-age) competence (see Schunk, 1987, p. 162).

To determine whether similarity in competence between model and observer is essential, it is important to distinguish different types of learning purposes. In situations involving the learning of skills or new behaviors, children tend to emulate competent peers. Modeled competence is necessary for students to learn correctly. Even coping models eventually demonstrate competence. But where social comparisons are used for self-evaluations, children are most interested in the performances of others who are similar in ability (Schunk, 1987, p. 162). Furthermore, all models, but especially models whom children view as similar in competence, may promote children's self-efficacy for learning observed skills (Schunk et al., 1987, p. 54).

The Similarity Hypothesis in Observational Learning to Write

In the present study we focused on writing argumentative texts, which was a new task for the eighth-grade participants. We assumed that the process of observation differs between weak and good students. When weaker students observe and evaluate a model, they rely to a lesser extent on knowledge about good writing. In contrast, good students, who already have access to a—albeit imperfect—knowledge base and a set of criteria (Hillocks, 1986), can apply this knowledge to the evaluation of the model.

When weak writers observe a competent and a noncompetent model, it is much easier for them to evaluate the weak model; they can use the performance of the better model as a reference. Besides, the performance of the observed weak model is probably more matched with the cognitive processes of a weak learner. The same can be argued for the observation of good models for good writers; the performance of a good model is probably more matched with the cognitive processes of a good learner. Furthermore, when good writers observe pairs of models, they are able to identify and comment on the qualities of the good model, using their own set of evaluation criteria.

This brings us to the similarity hypothesis, which is tested in the present study. This hypothesis states that weak writers learn more from focusing their observations on weak models and that good writers learn more from focusing on good models. In the study, participants were assigned to one of two observational-learning conditions or to a control condition. The two observational-learning groups observed the same pairs of models. The participants focused respectively on the noncompetent (weak) model (observation/weak-focus condition) or on the competent (good) model (observation/good-focus condition). Participants in the control group (direct-writing condition) performed the writing tasks themselves. The addition of the direct-writing condition enabled us to compare the relative strengths and weaknesses of two types of observational learning with a more traditional type of learning.

Method

Participants

The experiment involved 214 fourteen-year-old participants from nine eighth-grade classes in a Dutch city school. The achievement level of the participants ranged from low to high. Classes were composed heterogeneously in terms of achievement level and gender. The experiment was part of the regular lesson roster, therefore classes were randomly assigned to conditions. There were no differences in gender and age between conditions. In the direct-writing condition, 54% of the participants were girls. Mean age was 14.01 years. In the observation/weak-focus condition, 55% of the participants were girls. Mean age was 14.12 years. Finally, in the observation/good-focus condition, 50% of the participants were girls. Mean age was 14.15 years.

Design

The general design of the present study was a pretest—posttest control-group design (see Table 1). We implemented two experimental instruction sessions. In the first instruction session we used a CD-ROM to show the models, and in the second instruction session we used "live modeling." After each of these sessions, the effect on performance was measured.

By implementing two instruction sessions, we could test the similarity hypothesis in two circumstances: when a task is completely new (Instruction session 1) and when a task had already become a bit regular (Instruction session 2). Furthermore, implementing two variants of the instruction might enhance the validity of the independent variable, as it allows generalizing effects over different specimens of the independent variable.

Participation in the experiment took place in six sessions, each lasting approximately 30 to 60 min, conducted by the researcher and research assistants. Pretests were administered during the first two sessions. The third session was the first instruction session. Posttests were administered during the fourth session. The second instruction session was implemented in Session five; some posttests were administered in Session six (see Table 1). The sessions were distributed equally over a 5-week period.

Materials and Procedure

For this study, we developed a short course on an aspect of writing argumentative texts, consisting of two instructional sessions. Students learned to transform argumentation structures into short linear argumentative texts. This learning task was completely new for all of the participants. The course was adapted with permission from Couzijn (1995, 1999). Students applied the theory presented in short writing tasks (i.e., the direct-writing control condition) or in observation tasks. During the latter, participants observed pairs of models performing short writing tasks, which were the same as those the participants in the control condition performed. Participants in both observational-learning conditions observed the same pairs of models (five pairs in Instruction session 1; two pairs in Instruction session 2), in the same order; the only difference was that instructions in

Table 1
Research Design

Session	Content	Direct writing (n = 76)	Observation/weak focus (n = 74)	Observation/good focus (n = 64)
1	Pretest 1 (verbal intelligence)	X	X	X
2	Pretest 2 (identification of argumentation)	X	X	X
3	Instruction session 1	Writing texts	Observing two models on CD-ROM who are writing texts, focus on the weak model	Observing two models on CD-ROM who are writing texts, focus on the good model
4	Posttest 1 (writing)	X	X	X
5	Instruction session 2	Writing texts	Observing two "live" models who are writing texts, focus on the weak model	Observing two "live" models who are writing texts, focus on the good model
6	Posttest 2 (writing)	X	X	X

Note. X indicates that these tests were administered.

the observation/weak-focus condition required participants to reflect on the performance of the weak model, and those in the observation/good-focus condition required participants to reflect on the performance of the good model.

The first instruction session lasted about 60 min. Materials for this session consisted of a workbook and a CD-ROM program that was projected onto a large screen in front of the class. Participants were informed of what to do by means of on-screen messages (e.g., reading theory in the workbook, answering a mastery question, writing a short text, or observing two models). An on-screen timer indicated how much time was left for each activity; participants were alerted by short beeps when the time had almost elapsed. In pilot sessions, the required time for each activity and the difficulty of theory, mastery questions, and exercises were examined. The required time was calculated in such a way that it was possible to read each theoretical section once, but participants were allowed to glance backwards through the workbook, for instance to reread parts of the theoretical sections when answering a mastery question.

The participants' workbooks consisted of four main sequences (see Figure 1). In the first sequence, participants had to study a theoretical section. To stimulate active reading of the theory, all of the participants answered five mastery questions. In the second sequence, participants applied the theory in different types of exercises according to the condition to which they had been assigned. This meant that participants in the direct-writing condition wrote short argumentative texts based on the argumentation structures they had been given. Meanwhile, participants in the observation/weak-focus condition used a CD-ROM to observe pairs of videotaped models who were writing short argumentative texts on the basis of these argumentation structures. In two questions, they were instructed to focus their attention on the performance of the weak model. Participants in the observation/good-focus condition observed the same CD-ROM, with the same pairs of models writing the texts. But in this case they were instructed to focus their attention on the good model. By using videotaped models on CD-ROM, standardized presentation across participants was ensured.

After completing the exercises, the participants went on to the third sequence, which contained theory and mastery questions about subordinate argumentation, argumentation structures, and argumentative connectives. Finally, in the fourth sequence, they performed two writing tasks or an observation task, depending on the condition.

To provide equal learning time in all conditions, we had to add two extra writing tasks in the control condition, because observing two models writing a text and answering two questions required more time than to write a text. Therefore, participants in the direct-writing control condition performed seven writing tasks in this instruction session, and participants in both observational-learning conditions performed five observation tasks. In Appendix A, an overview of instructional time for each sequence is presented. The total learning time in the control condition and in both observational-learning conditions was almost equal.

The second instruction session lasted 30 min, starting with a summary of the theory presented in the previous session and some mastery questions. There were opportunities for participants to ask questions and to discuss their answers with their peers. The discussions were about the correct answers on the mastery questions and why an alternative answer could or could not be correct. For instance, one of the mastery questions was about the correct connective between standpoint and argument. One of the correct answers was the connective *because*. Some participants asked whether their answer, *thus*, was correct. That answer was incorrect because with the connective *thus*, the argument should be presented first, followed by the standpoint.

The different groups then performed the following activities: The participants in the direct-writing condition wrote two short argumentative texts based on the argumentation structures they had been given. After writing each text, there was a short discussion between the participants about which texts were possible. For instance, participants discussed which connective should be used to connect the subordinate argument with the main argument. This is an important issue because if one uses a coordinating connective such as *and*, the subordinate argument becomes coordinative instead of subordinate. This discussion was led by the researcher.

The participants in the observational-learning conditions performed two observation tasks. These tasks and sequencing were similar to the tasks in Instruction session 1. However, instead of watching models on CD-ROM, the observers watched live models (i.e., a researcher and a research assistant following a script). The use of a script here also ensured a standardized presentation across participants; in order to avoid a researcher-effect, the roles were randomly assigned. As in Instruction session 1, participants in both conditions observed the same pairs of models and were instructed to focus on the weak model (observation/weak-focus condition) or on the good model (observation/good-focus condition). After each observation task, the researcher led a short discussion between the participants on the possible answers to the two questions. They discussed why one model was not as good (observation/weak-focus condition) or was better (observation/good-focus condition) than the other model. For instance, in one of the tasks, the weaker model wrote the subordinated argument as a coordinated argument.

The first part of Instruction session 2 (summary and mastery questions) lasted 10 min, and the second part (performing two writing tasks or two observation tasks) lasted 20 min (10 min for each task).

Writing argumentative texts

Sequence I: Theory and five mastery questions about seven main concepts (all conditions)

- standpoint (opinion) (for instance, the music should not be so loud)
- argument (reason) (for instance, through the loudness I cannot work)
- argumentative texts (text in which one gives his standpoint and argumentation for his standpoint)
- singular argumentation (standpoint with only one argument)
- compound argumentation (standpoint with two or more arguments)
- argumentation structures (structure of standpoint and arguments)
- argumentative connectives (words to connect standpoint with arguments, for instance, "because")

Example mastery question: "What is the difference between singular and compound argumentation?"

↓

Sequence II: Applying the theory in five short writing tasks (direct writing) or four observation tasks with the focus on *weak* models (observation/weak focus) or focus on *good* models (observation/good focus). For examples of writing and observation tasks see Conditions section.

↓

Sequence III: Theory and four mastery questions about subordinate argumentation (argument which supports an argument), argumentation structures (structure of standpoint, arguments and subordinate arguments) and argumentative connectives (words to connect standpoint with arguments, and arguments with subordinate arguments, for instance, "because") (all conditions).

Example mastery question: "When is an argument a subordinate argument?"

↓

Sequence IV: Applying the theory in two short writing tasks (direct writing) or one observation task with the focus on the *weak* model (observation/weak focus) or focus on the *good* model (observation/good focus). For examples of writing and observation tasks see Conditions section.

Figure 1. Sequences of Instruction session 1 "writing argumentative texts."

Conditions

Participants in the direct-writing condition applied the theoretical material in writing tasks. They wrote short argumentative texts based on argumentation structures (see Figure 2).

Participants in both observational-learning conditions observed the same pairs of models (a competent and a noncompetent model) executing the same writing tasks. Because the models were thinking aloud while executing the writing task, the participants gained full insight into the writing process and into the task solution (i.e., the writing product) of the models. Besides, some models also gave the participants insight into their monitoring activities, demonstrating monitoring, evaluating, and reflecting on performance. The verbalizations of the models' thoughts were directed to (planning and checking) the writing processes, not to achievement beliefs. Our weak models did not verbalize achievement beliefs that reflected low self-efficacy and negative attitudes, and neither did our good models verbalize positive statements reflecting high self-efficacy.

The participants were stimulated to make notes when they were observing the models, as these might help them answering two questions after observing the two models. These two questions distinguished between the two observational-learning conditions.

Participants in the observation/weak-focus condition were instructed to focus on the weak model by answering the questions "Which model did less well?" and "Explain briefly what this (less good) model did worse." When answering these questions, the participants were stimulated to evaluate the writing processes and the resulting products, and also to reflect on them. Figure 3 is an example of an observation task for participants in the observation/weak-focus condition, drawn from Instruction session 1. A possible answer on the questions in the Figure 3 observation task would be, "Model one did worse because he wrote first all arguments and then the arguments again with the subordinate arguments. It took too long!"

Participants in the observation/good-focus condition observed the same pairs of models writing argumentative texts, but answered different question, "Which model did well?" and "Explain briefly what this (good) model did well." One of the participants in the observation/good-focus condition wrote in his answer on the task equivalent to the one in Figure 3: "Model two did well because she used 'firstly,' 'secondly,' and 'moreover,' and she connected the arguments directly with the subordinate arguments."

The answers of the participants on the observation tasks were scored for accuracy of implementation. We found that in both conditions, participants focused on the intended model five of seven times.

Testing Materials

To indicate which participants were weak and which were good, we measured their aptitude. To ensure that aptitude would not be distorted by the effects of the instruction sessions, we performed this measurement before the instruction sessions took place. We measured aptitude by two

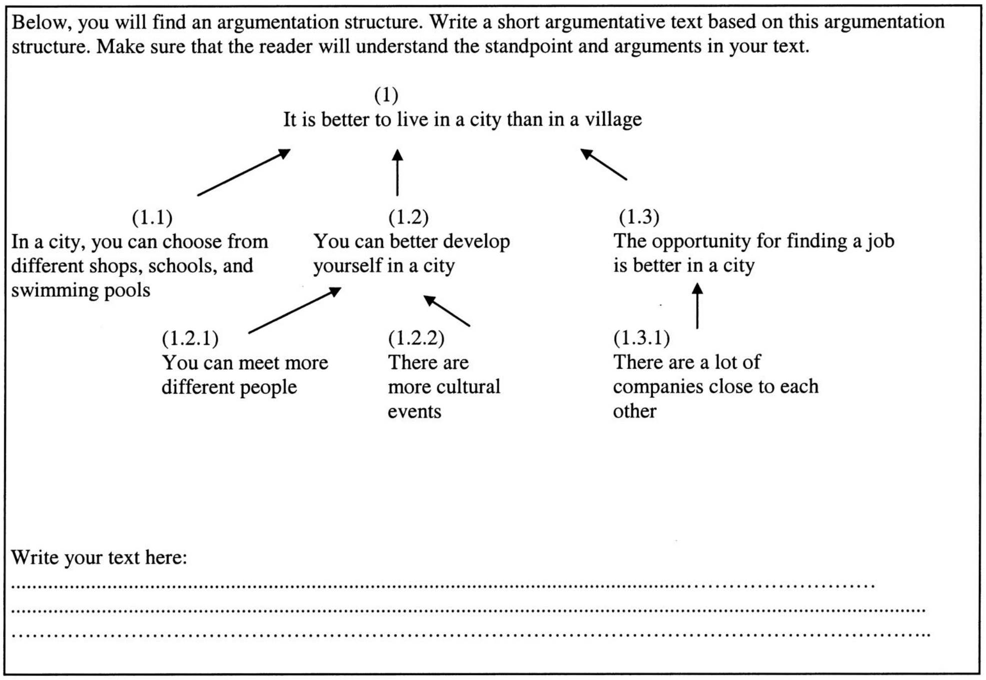


Figure 2. Writing task for participants in direct-writing condition.

intelligence tests: Cognition of Meaningful Units Word List (DAT, 1984) and Verbal Analogies (DAT, 1984). We chose these particular tests because the analysis of argumentation (an important skill when learning to write argumentative texts) can be interpreted as an ability to discern abstract relationships between verbal units (Oostdam, 1991). The tests scores were summed, which resulted in a reasonably reliable measurement (Cronbach's $\alpha = .69$). The mean[1] and standard deviation for each condition are in Table 2. An example of this measurement of aptitude is in Table 3.

To control for a priori differences between conditions on the writing of argumentative texts, all of the participants performed a pretest that measured their skill in identifying argumentation, a skill that is very important to the writing of argumentative texts. The reliability (Cronbach's alpha) of this pretest is .72. The mean and standard deviation for each condition are in Table 2. An example of an item from this pretest is in Table 3. There were no significant initial differences in mean scores between conditions, neither on the measurement of aptitude, $\chi^2(3) = 1.49$, $p = .68$, nor on the pretest on the writing of argumentative texts, $\chi^2(3) = 5.01$, $p = .17$.

Writing performance of each participant after Instruction session 1 (Posttest 1) and Instruction session 2 (Posttest 2) was measured by writing tests in which participants wrote five argumentative texts based on five argumentation structures. Participants had 40 min to write their texts in Posttest 1, and 25 min to write their texts in Posttest 2. There was sufficient time to complete the tests; 88% of the participants had no missing items on Posttest 1, and 82% had no missing items on Posttest 2. An example of the type of the assignment can be found in Figure 2; the type of test was the same as the writing tasks undertaken by participants in the direct-writing condition.

The tests were coded to measure whether the participants correctly identified the standpoint, arguments, and subordinate arguments in the texts. This measurement was performed by examining whether the participant used appropriate connectives to link standpoints with arguments, and arguments with other arguments or subordinate arguments. For instance, if a standpoint is written first and a participant connected the standpoint and the argument with the connective *thus,* this is coded as incorrect. For the correct code, the participant had to use the connective *because.* The completeness aspect was measured by coding missing items as incorrect. We coded strictly; the participant had to use explicit connectives. If he or she connected a standpoint with an argument by using a comma, it was coded as incorrect.

The tests given after the two instruction sessions were of a similar type, the only difference being one of content. To avoid possible effects of content, a fully counterbalanced design was used: One half of the participants received Writing test A as Posttest 1, and Writing test B as Posttest 2. The other half received Writing test B as Posttest 1 and Writing test A as Posttest 2. The items from text four and five (the more difficult texts) of both tests were used for measuring writing performance. The reliability (Cronbach's alpha) of Posttest 1 was .68, and that of Posttest 2 was .73.

Analyses

Suppose that Y_{ij} is the score of participant j ($j = 1, 2, 3, \ldots, j$) at occasion i ($i = 1, 2$) in the direct-writing condition. Let T_{1ij} and T_{2ij} be dummy variables that are turned on (equal 1) at Occasions 1 and 2, respectively, and otherwise are turned off (equal 0). Furthermore, APT_{0j} stands for the aptitude of individual j and $PERF_{1j}$ for previous performance of this individual. The model[2] to be analyzed (the aptitude instruction model) for this condition could thus be written as

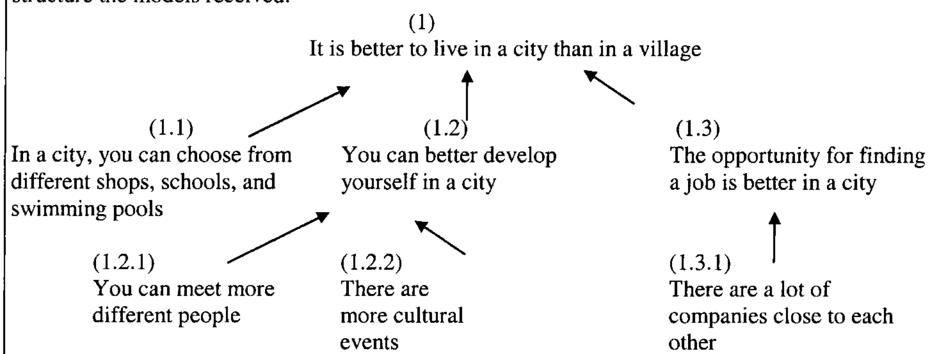
$$Y_{ij} = T_{1ij}(\beta_1 + \beta_2 \times APT_{0j} + u_{10j});$$
$$T_{2ij}(\beta_3 + \beta_4 \times APT_{0j} + \beta_5 \times PERF_{1j} + u_{20j}). \quad (1)$$

In this model, β_1 and β_3 indicate the intercept of first and second measurement occasion, respectively. The influence of aptitude on both measurement occasions can be evaluated by means of β_2 and β_4; these parameters are the regressions of both posttest scores on aptitude scores. The fifth regression weight (β_5) was introduced to the model in order to

[1] For all tests, a normalization transformation was applied to fulfill the assumptions required for the analyses.

[2] We did not specify classroom level in the model because the unit of analysis is the individual: The intervention was applied to individuals, not to classes. Hypotheses about differential effects on classes—all composed as mixed ability classes—were not plausible. Furthermore, "only" nine classes were involved. The class-level coefficients in the model would never reach significance.

In a little while, you are going to watch on CD-ROM videotape recordings. You will see two models writing a short argumentative text based on the following argumentation structure. The models had to make sure that the reader will understand the standpoint and arguments in their text. Below, you will find the argumentation structure the models received.



After watching the models, you have to answer the following questions:

1. Which model did less well?
2. Explain briefly what this (less good) model did worse

When you have observed both models, you may advance to the next page

Make your notes here, when you observe the models:
....
....
....
....

(next page)

You saw two models doing the assignment. They wrote the following texts:

Model 1
It is better to live in a city than in a village because in a city you can choose from different shops, schools, and swimming pools, you can better develop yourself in a city and the opportunity for finding a job is better in a city. Because you can meet more different people and there are more cultural events, you can better develop yourself. The opportunity for finding a job is better because there are a lot of companies close to each other. Therefore it is better to live in a city.

Model 2
It is better to live in a city than in a village because firstly in a city you can choose from different shops, schools, and swimming pools. Secondly you can better develop yourself in a city because you can meet more different people and there are more cultural events. Moreover, the opportunity for finding a job is better in a city because there are a lot of companies close to each other.

1. Which model did less well? Model ....
2. Explain briefly what this (less good) model did worse.

Model ...... did worse,
because...
....
....

Figure 3. Observation task for participants in the observation/weak-focus condition.

take into account differences due to the first experimental instruction session. In a more traditional analysis, this parameter would be called a covariate. The last two parameters, u_{10j} and u_{20j}, are residual scores that index the differences between the mean score and the writing score of participant j in the first and second posttest, respectively. The residuals are assumed to be normally distributed with a variance of S^2u_1 and S^2u_2.

This model, which is given for one condition, can easily be extended to the case at hand, in which there are three conditions to be distinguished. In this case (three conditions multiplied by 5), 15 regression weights need to be estimated.

The effect of individual variables can easily be tested, because the ratio of the parameter's estimate and its standard error is student-*t* distributed.

Table 2
Number of Items, Maximum Score, Means, and Standard Deviations for Each Condition

Measurement	No. of items	Maximum score	Direct writing *M*	Direct writing *SD*	Observation/ weak focus *M*	Observation/ weak focus *SD*	Observation/ good focus *M*	Observation/ good focus *SD*
Aptitude	22	22	5.36	2.86	5.94	3.06	6.13	3.61
Identification of argumentation	8	8	2.88	2.08	3.31	1.80	3.57	1.90

Differences between conditions can be tested by means of a contrast procedure (Goldstein, 1995).

In this model, all explanatory variables (i.e., aptitude and performance) are centered around the grand means. Thus, if differences between conditions in the intercept of the first measurement occasion are assessed, the writing score of participants with a mean aptitude score differs between conditions. For the intercept of the second measurement occasion, the exact interpretation is the mean of participants' writing scores with a mean aptitude score as well as a mean performance score.

Differences between conditions in effects of aptitude can be assessed on both measurement occasions. This allows for a direct testing of our similarity hypothesis. If participants benefit from a similarity with the models, a relatively more positive regression weight is expected in the observation/good-focus condition than in the observation/weak-focus condition.

Results

Table 4 shows all estimates for the parameters in our model. The first two rows concern Posttest 1, and report the intercepts for Posttest 1 and the regression weights of Posttest 1 on aptitude. The other rows concern Posttest 2. Next to the intercepts for Posttest 2, two regressions are reported: the regression weights of Posttest 2 on aptitude and the regression weights of Posttest 2 on Posttest 1. In Appendix B, the means and standard deviations for both writing measurements in each condition are reported.

The regression of Posttest 1 on aptitude was not statistically significant ($p > .05$) in the observation/weak-focus condition (see Table 4). In other words, the aptitude of these participants made no difference to their writing scores for Posttest 1. However, in the two other conditions (i.e., direct writing and observation/good focus) the regression was statistically significant ($p < .05$) and did not differ between these two conditions, $\chi^2(1) = 1.08$, $p = .30$. In these conditions, aptitude influenced writing performance in a similar fashion: Participants with a higher aptitude benefited more from the instruction. Thus, the effect of Instruction session 1 in both the direct-writing and the observation/good-focus condition differed according to the level of aptitude. Figure 4 is a schematic representation of the interactions.

When we tested the intercept scores for Posttest 1 (i.e., the mean writing scores for Posttest 1 for participants with a medium aptitude), we found no significant differences between conditions, for all comparisons $\chi^2(1) < .60$, $p > .44$. Thus, for a participant with a medium aptitude, condition did not affect performance. However, it is better for participants with a low aptitude to be assigned to the observation/weak-focus condition, and for participants with a high aptitude to be assigned to direct-writing or the observation/good-focus condition (see also Figure 4).

The regression of Posttest 2 on aptitude was not statistically significant in the direct-writing and observation/weak-focus condition ($p > .05$; Table 4). Thus, the aptitude of these participants made no difference to their writing scores for Posttest 2. In the observation/good-focus condition, the regression of Posttest 2 on aptitude was statistically significant ($p < .05$). Aptitude influenced the writing scores: Participants with a higher aptitude benefited more from Instruction Session 2 (see Figure 5).

It should be noted that the regression of Posttest 2 on aptitude was estimated simultaneously with the regression of Posttest 2 on Posttest 1. This regression of Posttest 2 on Posttest 1 was statistically significant in all conditions ($p < .05$) and did not differ between conditions, for all comparisons $\chi^2(1) < 1.63$, $p > .20$ (Table 4). This means that the effects of aptitude on Posttest 2 can be maintained for each level of performance. Thus, considering all effects on Posttest 2, we can conclude that, for all conditions, there is an interaction between writing performance at Posttest 1, and instruction in Session 2. Writing performance at Posttest 1 influences writing scores for Posttest 2 in all conditions: Participants

Table 3
Examples of Aptitude and Identification of Argumentation

Measurement	Example pretest item
Aptitude	Make a complete and logical sentence: . . . is to hear as blind is to . . . . (a) Ear–light (b) Deaf–see (c) Deaf–dark (d) Sound–eye (e) Noise–watch
Identification of argumentation	Underline the argument(s) in the following text: Your brother was not allowed to go on his own on vacation when he was 16. Your sister was not allowed too. You are also not allowed to go on your own on vacation.

Table 4
Estimates for All Parameters in the Aptitude Instruction Model

	β (*SE*)		
Parameter	Direct writing	Observation/weak focus	Observation/good focus
Intercept for Posttest 1	6.13 (0.22)	6.20 (0.23)	6.38 (0.24)
Regression of Posttest 1 on aptitude[a]	0.29 (0.08)	*ns*	0.18 (0.07)
Intercept for Posttest 2	7.06 (0.21)	6.63 (0.24)	6.47 (0.25)
Regression of Posttest 2 on aptitude[a]	*ns*	*ns*	0.14 (0.07)
Regression of Posttest 2 on Posttest 1[a]	0.55 (0.10)	0.35 (0.12)	0.47 (0.14)

Note. Standard errors are in parentheses.
[a]Variables are centered around the mean.

with a good writing performance learned more than participants whose writing performance was not as good. On the other hand, the interaction between aptitude and instruction in Session 2 differs. Whereas there was no interaction between aptitude and instruction for the direct-writing and observation/weak-focus conditions, an effect of aptitude on writing scores for Posttest 2 was found for the observation/good-focus condition. Thus, in the latter, a good writing performance is not enough to gain higher writing scores for Posttest 2; to benefit from this performance, participants also have to be smart.

When we tested the intercept scores for Posttest 2 (i.e., the mean writing scores for Posttest 2 for participants with a medium aptitude and a medium writing performance at Posttest 1), we found no significant differences between conditions, for all comparisons $\chi^2(1) < 3.44$, $p > .06$. Thus, for a participant with a medium aptitude and a medium writing performance condition did not affect performance. However, it is better for participants with a low aptitude and low performance to be assigned to the observation/weak-focus or direct-writing condition and for participants with a high aptitude and high performance to be assigned to the observation/good-focus condition (see Figure 5).

Discussion

Results confirmed the similarity hypothesis; both instruction sessions showed that model–observer similarity in competence facilitated learning. Furthermore, results after Instruction session 1 showed that weak students benefited more from observational learning (focusing on weak models) than by performing writing tasks. We assume that they profited from observational learning because their cognitive effort is shifted from executing writing tasks to learning from writing processes of others (Braaksma, van

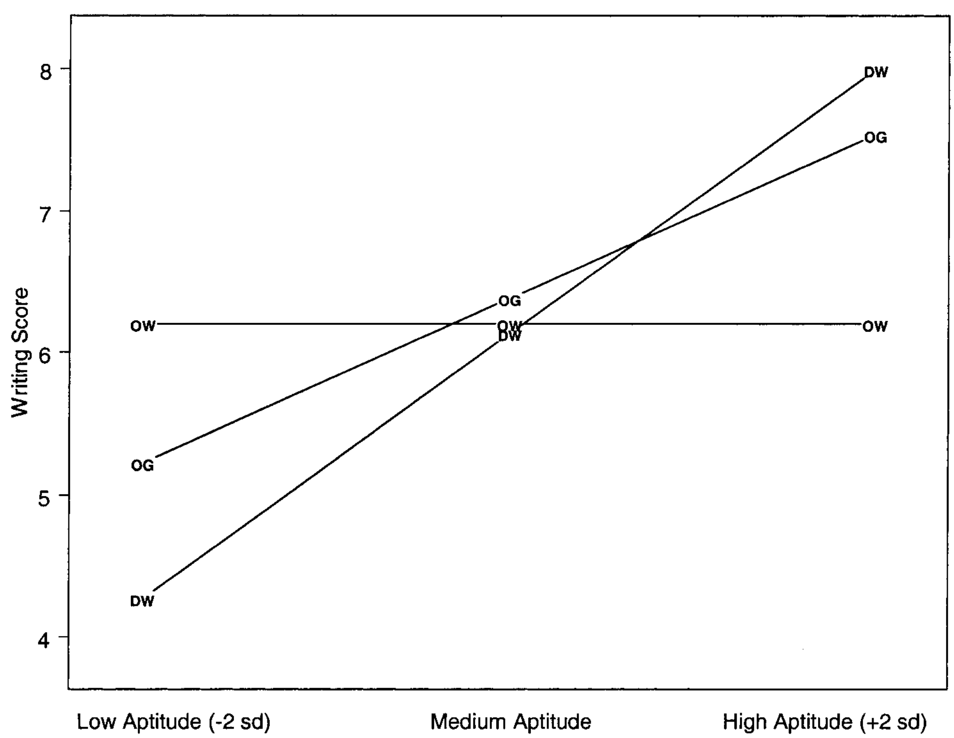


Figure 4. Effects of instruction for different levels of aptitude (Posttest 1). OW = observation/weak-focus condition; OG = observation/good-focus condition; DW = direct-writing condition.

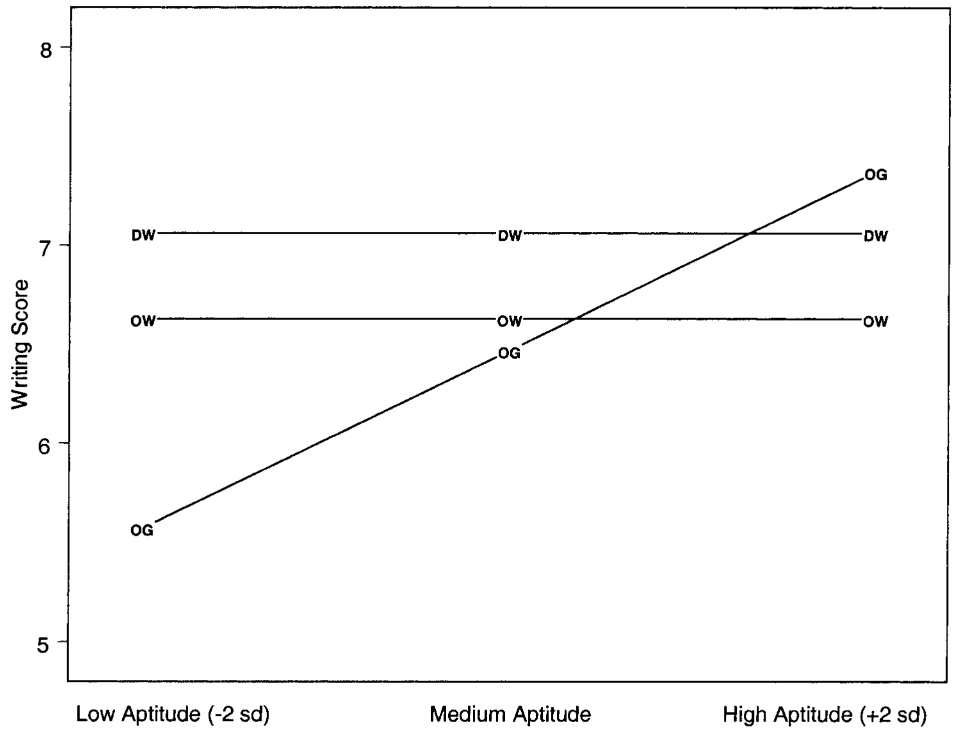


Figure 5. Effects of instruction for different levels of aptitude (Posttest 2), conditional on writing performance at Posttest 1. DW = direct-writing condition; OW = observation/weak-focus condition; OG = observation/good-focus condition.

den Bergh, Rijlaarsdam, & Couzijn, 2001; Couzijn, 1999; Rijlaarsdam & Couzijn, 2000a, 2000b). They can thus focus on the learning task, providing themselves with a learning opportunity to acquire new understanding about writing. Good students benefited not only from observational learning (focusing on good models) but also from performing writing tasks. They are probably able to divide their attention between writing task and learning task, and thus generate enough input for their learning by evaluating their own performance. Instruction session 2 showed that weak participants benefited from performing writing tasks as much as they profited from focusing their observations on weak models' writing. Perhaps familiarity with the task played a role here; because they had already experienced successes with the tasks, they had a chance to build a knowledge base about good writing and, thus, became better equipped to learn from executing the writing task.

The writing tasks selected for our study were constrained. What the participants were called upon to do (and what the models they observe were doing) was order and connect the ideas we gave them. The participants did not have to come up with new ideas and compose thoughts, as in the larger, less structured writing tasks that have been used in other writing studies in which modeling was a key factor (e.g., Couzijn, 1999). In future studies on the effects of model–observer similarity on writing, we thus recommend other, less structured, and larger writing tasks.

Until this study, the similarity claim was restricted to a group of weak students (Schunk et al., 1987). When a group of average performers was studied, no effects of similarity between model and observer were found (Schunk & Hanson, 1989a). Our study, however, involved a heterogeneous group of participants (i.e., participants with mixed abilities), which enabled us to extend the similarity hypothesis to good learners.

Another extension relates to the operationalization of similarity. In most studies, participants in the good-model condition observed good models, and participants in the weak-model condition observed weak models. In our case, however, all of the participants were exposed to pairs of a good and a weak model. The only difference between conditions was the instruction to focus their reflection on the weak or on the good model. Observers watched both models at work and compared them instead of observing just one type of model. Thus, the similarity claim also applies when the exposure to models is the same but the instruction is similarity inducing.

Because of the different operationalization of similarity, we made other choices concerning the execution of the models and the dependent variables. Although our models did verbalize their thoughts, they focused on the execution of the writing process, not on achievement beliefs, and did not reflect (high and low) self-efficacy. Because of that different focus, we did not include self-efficacy measures as dependent variables. However, in future studies it might be advisable to include these measures, because self-efficacy is hypothesized to influence choice of activities, effort expenditure, persistence, and achievement (Bandura, 1986, 1997; Schunk, 1998). Perhaps the present outcomes are mediated by changes in participants' self-efficacy.

Finally, interaction effects played a central role in our study. We examined the interaction between characteristics of the participants and characteristics of instructions. With interaction effects, one can examine which learning environment is the most effective for which student. With this idea in mind, our results are in line with Cronbach (1957).

> In general, unless one treatment is clearly best for everyone, treatments should be differentiated in such a way as to maximize their interaction with aptitude variables. Conversely, persons should be allocated on the basis of those aptitudes which have the greatest interaction with treatment variables. (p. 681)

Observational learning may not be the best learning environment for all students, and not all types of observational learning are equally effective for different types of students.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy. The exercise of control.* New York: Freeman.

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, 1,* 33–48.

Couzijn, M. J. (1995). *Observation of writing and reading activities. Effects on learning and transfer.* Unpublished dissertation, University of Amsterdam, Amsterdam, the Netherlands.

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on learning and transfer. *Learning and Instruction, 2,* 109–142.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

DAT. (1984). *Differentiële Aanleg Test* [Differential Aptitude Test]. Lisse, the Netherlands: Zwets & Zeitlinger.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Edward Arnold.

Graham S., & Harris, K. R. (1994). The role and development of self-regulation in the writing process. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 203–228). Hillsdale, NJ: Erlbaum.

Graham, S., Harris, K. R., MacArthur, C., & Schwartz, S. (1998). Writing instruction. In B. Y. L. Wong (Ed.), *Learning about learning disabilities* (2nd ed., pp. 391–421). San Diego, CA: Academic Press.

Graham, S., Harris, K. R., & Troia, G. A. (1998). Writing and self-regulation: Cases from the self-regulated strategy development model. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning, from teaching to self-reflective practice* (pp. 20–41). New York: Guilford Press.

Hillocks, G. (1986). *Research on written composition: New directions for teaching.* Urbana, IL: National Institute of Education, ERIC Clearinghouse on Reading and Communication Skills.

Oostdam, R. (1991). *Argumentatie in de peiling* [Assessing argumentation]. Unpublished dissertation, University of Amsterdam, Amsterdam, the Netherlands.

Rijlaarsdam, G., & Couzijn, M. (2000a). Stimulating awareness of learning in the writing curriculum. In G. Rijlaarsdam & E. Espéret (Series Eds.) & A. Camps & M. Milian (Vol. Eds.), *Studies in writing: Vol. 6. Metalinguistic activity in learning to write* (pp. 167–202). Amsterdam: Amsterdam University Press.

Rijlaarsdam, G., & Couzijn, M. (2000b). Writing and learning to write: A double challenge. In R. J. Simons, J. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 157–189). Dordrecht, the Netherlands: Kluwer Academic.

Rosenthal, T. L., & Zimmerman, B. J. (1978). *Social learning and cognition.* New York: Academic Press.

Schriver, K. A. (1992). Teaching writers to anticipate readers' needs: A classroom-evaluated pedagogy. *Written Communication, 9,* 179–208.

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research, 57,* 149–174.

Schunk, D. H. (1991). *Learning theories: An educational perspective.* New York: Merrill.

Schunk, D. H. (1995, March–April). *Social origins of self-regulatory competence: The role of observational learning through peer modeling.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Indianapolis, IN.

Schunk, D. H. (1998). Teaching elementary students to self-regulate practice of mathematical skills with modeling. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning, from teaching to self-reflective practice* (pp. 137–159). New York: Guilford Press.

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77,* 313–322.

Schunk, D. H., & Hanson, A. R. (1989a). Influence of peer-model attributes on children's beliefs and learning. *Journal of Educational Psychology, 81,* 431–434.

Schunk, D. H., & Hanson, A. R. (1989b). Self-modeling and children's cognitive skill learning. *Journal of Educational Psychology, 81,* 155–163.

Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79,* 54–61.

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist, 32,* 195–208.

Sonnenschein, S., & Whitehurst, G. J. (1983). Training referential communication skills: The limits of success. *Journal of Experimental Child Psychology, 35,* 426–436.

Sonnenschein, S., & Whitehurst, G. J. (1984). Developing referential communication: A hierarchy of skills. *Child Development, 55,* 1936–1945.

Appendix A

Overview of Instructional Time (in Minutes) for Each Sequence in Each Condition

Sequence	Direct writing	Observation/ weak focus	Observation/ good focus
I (theory and mastery questions)	14.5	14.5	14.5
II (writing or observation tasks)	21.0	22.5	22.5
III (theory and mastery questions)	11.5	11.5	11.5
IV (writing or observation tasks)	12.0	13.0	13.0
Total learning time	59.0	61.5	61.5

Appendix B

Means and Standard Deviations for Both Writing Measurements in Each Condition

	Posttest 1		Posttest 2	
Condition	*M*	*SD*	*M*	*SD*
Direct writing	6.04	2.09	7.09	2.10
Observation/weak focus	6.19	1.83	6.65	1.79
Observation/good focus	6.43	1.74	6.56	1.88

Received April 6, 2001
Revision received August 9, 2001
Accepted August 10, 2001 ■