Online Tutoring as a Pivotal Quality of Web-Based Early Literacy Programs

Cornelia A. T. Kegel and Adriana G. Bus
Leiden University

In this randomized controlled trial, 312 low-socioeconomic-status children (\(M_{\text{age}} = 52.9\) months, \(SD = 3.2\)) from 15 Dutch schools participated. Children in the intervention condition played early literacy games via the intelligent tutoring system Living Letters. Control children played a nonliteracy computer game. At the beginning of each intervention session, children received instruction from computer characters about how to play the game. While playing the game, half of the children in the intervention group received individualized feedback, which included oral corrections and cues from a computer tutor. The other half of the children received no individualized feedback. On average the intervention comprised 11 sessions (approximately 110 min). A main finding was that children’s code-related skills increased as a result of the Living Letters program but only when the program included a computer tutor that gave oral feedback to children’s correct responses and errors. Children with underdeveloped inhibitory control scored disproportionately low in a computer environment without tutoring.

Keywords: computer intervention, early literacy intervention, inhibitory control, online tutoring, randomized controlled trial

Before formal reading education begins, children acquire knowledge about code-related skills through activities such as parent–child book sharing (Bus, Van IJzendoorn, & Pellegrini, 1995; Mol & Bus, 2011) and joint writing activities (Levin & Aram, 2004). Because literacy experiences in low-socioeconomic (SES) families are often sparse, children from these families may enter school with less well developed code-related skills compared with peers from middle- to high-SES families (Shonkoff & Phillips, 2000; Stipek & Ryan, 1997) and consequently may be less successful in the first grades (Byrne, Fielding-Barnsley, & Ashley, 2000; Silva & Alves-Martins, 2002; Snider, 1995).

Attempts have been made to level initial differences in entry-level reading skills by exposing young children to special programs that promote code-related knowledge resulting in moderate effect sizes (e.g., Bus & van IJzendoorn, 1999; Ehri et al., 2001). Studies have considered whether computer interventions can provide similar instruction and practice, but overall efficacy of computer-assisted instruction (CAI) appears to be low (\(d = 0.19\)), according to a meta-analysis of 50 different experimental studies (Blok, Oostdam, Otter, & Overmaat, 2002). It is possible that tutoring in computer programs is not present to the same extent as it is in teacher-led interventions. We therefore hypothesized that computer programs might be more effective when they not only provide feedback about the correctness of the response—as most programs in the Blok et al. (2002) study do—but also provide explanations and suggestions to help children improve their responses. In this study we compare a program that provides only rudimentary feedback about the correctness of the response with a program that provides explanations and suggestions modeled on human tutors (Anderson, Boyle, & Reiser, 1985; Graesser, Conley, & Olney, in press; Van der Kooy-Hofland, Kegel, & Bus, 2011). We present the results of a randomized controlled trial of an educational computer treatment with and without availability of a built-in computer tutor that confirms correct responses and explains why the responses are correct or supplies suggestions to improve children’s responses.

As the number of computers in schools is now about one computer per five children (Kennisnet, 2010), the availability of educational software for young children has improved. Programs designed to teach core skills are also more available, although they are not yet used on a regular basis in classroom settings. Programs such as Number Race (Wilson, Dehaene, Dubois, & Fayol, 2009), Daisy Quest (Lonigan et al., 2003), and GraphoGame (Saine, Lerkkanen, Ahonen, Tolvanen, & Lytinen, 2011) are CAI programs focused on the teaching of basis numerical or phonological skills in a gamelike setting with only rudimentary feedback on the correctness of responses. In GraphoGame, for example, balloons showing correct answers turn green after they are chosen with clicks. However, in none of these programs is an explanation provided to children regarding why their answers are correct or how they might be improved, as is common practice in intelligent tutoring systems (ITSs).

What adults expect from children while playing computer games can be different from what children actually do (L. Labbo, ...
personal communication, October 11, 2004). The challenge for education is, therefore, to build programs that enhance learning but prevent children from mainly focusing on the fun part of the games and responding without reflection (Bodrova & Leong, 2006). In the Dutch ITS Living Letters (Van der Kooy-Hofland et al., 2011), an online tutor offers hints and corrections that are intended to focus students on target problems and aid them in solving those (Anderson et al., 1985; Graesser et al., in press). The research literature has indicated that tutoring is most effective when it immediately follows a response (Corbett & Anderson, 2001) and is personalized, meaning that help is adjusted to characteristics of the users or to the users’ interaction with the system (Vasilyeva, 2007). Living Letters builds on these general principles by providing three sorts of responses immediately following children’s reply: (1) repeating the instruction when children, on their first attempt to solve the game, just pick out an incorrect answer; (2) providing cues if they fail the same task once more; and (3) verbalizing the correct answer (at the end of each game, after the children found the correct solution themselves or after the online tutor modeled the answer, the program verbalizes how the correct solution can be found next time they encounter similar problems). The program thus provides not only feedback on the accuracy of answers, but it also offers oral cues to correct and optimize children’s responses (Fisch, 2005; Vasilyeva, Puuronen, Pechenizkiy, & Räsänen, 2007; Wild, 2009). We examined whether the tutor element in Living Letters increases the beneficial effects of the program and is worthwhile to consider when designing new games.

The original program Living Letters consists of a series of games designed for young children not yet demonstrating an awareness of the letter–sound relationship in an alphabetical language and aims to stimulate children to combine their understanding of how a familiar word, for example their name, looks with knowledge of how it sounds (Both-de Vries & Bus, 2008, 2010; Molfese, Beswick, Molnar, & Jacobi-Vessels, 2006). The program draws on surface perceptual knowledge of children’s names. Most young children develop this knowledge naturally when they encounter their name on personal belongings such as mugs and artwork (Levin, Both-de Vries, Aram, & Bus, 2005; Levin & Bus, 2003). The program stimulates the basic, but indispensable, understanding that letters in the name can be heard in the name’s spoken counterpart. The program’s instructional framework is modeled on how caregivers promote the development of letter–sound knowledge with the name as a starting point (Levin & Aram, 2004; Molfese et al., 2006). Analogous to children’s activities in daily life, the program emphasizes the following three successive skill areas: (1) recognizing their name in print, (2) associating the initial name letter with its sound, and (3) identifying the sound of the initial name letter in other words (Both-de Vries & Bus, 2010). A previous study of Living Letters revealed both short- and long-term effects of the program for a sample of low achievers at kindergarten age (Van der Kooy-Hofland et al., 2011). Children in the Living Letters group outperformed control children on early literacy tests completed directly after the program, as well as on word reading tests at the end of second grade.

In the current study, we focused specifically on the importance of the tutoring component of the Living Letters program. We therefore created a version of Living Letters without a tutor (a revised program) in addition to the original program with a tutor. In both versions of the program, games and instructions were the same, and children received an identical number of trials and repetitions. The two programs differed only in the presence of an online tutor to provide oral feedback to children’s responses.

A second aim of this study was to test whether a subsample of children with less-developed inhibitory control is more susceptible to the presence of an online tutor than are the rest. Previous research has demonstrated that children with regulatory skills in the normal range benefit more from a literate environment (e.g., Davidse, De Jong, Bus, Huijbrechts, & Swaab, 2011) including computer games (Kegel, Van der Kooy-Hofland, & Bus, 2009). Working memory, one aspect of regulatory skills, may be less vital for learning of our computer intervention because the relatively short and simple games do not strongly appeal to retention and manipulation of information (Diamond, Barnett, Thomas, & Munro, 2007). Inhibitory control—another component of self-regulation, which involves withholding or restraining a motor response in favor of a potentially less-dominant but more-adaptive response—may be necessary to stay on task and follow the rules of the computer games (Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009; Diamond et al., 2007; Lonigan et al., 1999).

In particular, children with poor inhibitory control may be disadvantaged by a program that lacks an online tutor. These children are easily distracted, and without a program that orally corrects and confirms responses and offers suggestions for improvement of problem solutions they may react randomly to computer assignments, which may in turn result in low achievement. Poor inhibitory control combined with a less-supportive environment may thus create a “dual risk” for widening the knowledge gap (Belsky, Bakermans-Kranenburg, & Van Ijzendoorn, 2007). Children scoring in the normal range, however, may suffer less when a program lacks an online tutor. These children may be less dependent on program qualities, because their inhibitory control may compensate for a less-optimal environment. It seems reasonable, therefore, to hypothesize that children with low inhibitory control may be adversely affected by a computer environment that lacks oral support designed to aid them in problem solving. Poor inhibitory control may not hinder learning in a “positive” environment but may do so in a “negative” environment (Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Blair & Razza, 2007; McClelland et al., 2007).

This Study

If a tutor offering oral feedback, hints, and explanations is important in computer-assisted learning (Vasilyeva, 2007), then Living Letters may not be as effective without online tutoring, even if the assignments, instructions, and number of task repetitions remain the same (Meyer et al., 2010). In particular, children with low inhibitory control may be at dual risk when a computer program does not provide online tutoring. When a program neither corrects impulsive responses nor offers suggestions for finding the correct solutions after errors, it may reward impulsive reactions and therefore enhance a tendency to respond without reflection. Children are at double risk not to benefit from the program when the environment does not reinforce their regulatory skills. An earlier study (Kegel et al., 2009) supported the hypothesis that weak regulatory skills elicit random computer behavior, thus limiting learning from the ITS Living Letters. However, studies so far
have not examined whether regulatory skills moderate the effects of computer instruction, especially when the program fails to offer personalized, oral support.

The study addressed the following research questions:

1. Can an intelligent tutoring system promote young (low-SES) children's foundational code-related knowledge? Living Letters, a computer program for preschoolers with delays in school-entry skills, may foster the development of code-related knowledge.

2. Is an online tutor that provides immediate, personalized oral feedback, explanations, and hints a vital component of a computer program designed to promote preschoolers' foundational code-related knowledge? The ITS Living Letters, with a built-in computer tutor, may be more effective than a CAI program that includes the same assignments, instructions, and number of task repetitions but provides only subtle feedback on correctness of the answer.

3. Does a tutor providing immediate, personalized oral feedback, explanations, and hints affect the quality of children's responses, and does the children's computer behavior predict gains? An online tutor may stimulate children to respond more thoughtfully, which may result in fewer errors in assignments and better posttest scores.

4. Do children's regulatory skills moderate program effects? Working memory may not moderate program effects, because the tasks are simple; however, underdeveloped inhibitory control may level the efficacy of computer activities, especially when there is no tutor to correct behavior. As a result, children may be at dual risk, especially when poor inhibitory control is not leveled by a compensatory computer environment.

Method

Participants

Participants were 312 kindergartners (60% male) from 15 Dutch schools in Rotterdam, Leiden, and the surrounding areas. Schools were selected for inclusion if they served large numbers of low-SES families and agreed to participate. For 70% of the mothers in our sample, their highest level of education was senior secondary vocational education (about 13 years of education, excluding pre-kindergarten). Children who were about 4 years old (M = 52.9 months, SD = 3.2) at the beginning of the year in which the intervention was carried out, and who spoke Dutch as their first language, qualified for participation in the experiment. Parental consent was obtained, with a positive response rate of 91%. Each school received 1,000 euros (about $1,350) for participation in the experiment.

Study Design

A randomized controlled trial design was used to examine the effects of the ITS Living Letters. Two Living Letters intervention conditions were created, one with a tutor (LL-Tutor) and one without online tutoring (LL-NoTutor). The first program is the original program examined in earlier studies (Kegel et al., 2009; Van der Kooy-Hofland et al., 2011). Two control groups were assigned to another computer program (Clever Together). In this study, these two control groups were reported as one condition because there were no between-group differences in pre- and posttest scores on outcome measures. Eligible pupils were randomly assigned by the main researcher to a condition stratified for school, gender, and children's level of regulatory skills (knock and tap) on a pretest (see Table 1).

Programs

Living Letters. The ITS Living Letters—designed by a team of computer experts, designers, and experts in the field of education and available for schools and parents via subscription (www.Bereslim.nl)—aims to train foundational code-related skills. The child’s name or another familiar word (i.e., mama [mommy]; Levin, Shatil-Carmon, & Asif-Rave, 2006) is used to draw attention to the relationship between letters in a name’s visual form and phonemes in the spoken name. Because a name is usually the first word that young children can read and write, children received the program version with their name unless the name’s spelling was inconsistent with Dutch orthography (e.g., Chris or Joey). In those cases (22% of the sample), the program used mama as the target word (Both-de Vries & Bus, 2008, 2010).

The computer program begins with 20 games in which children practice finding their name or mama between other signs and words (see Figures 1a and 1b), followed by 10 games targeting the sound of the first letter of their name or mama (e.g., “Which one of the letters (e.g., u, s, x) is the first letter of Sim?”; see Figure 1c) and 10 games in which children are given the task to identify pictures that start with or contain the first letter of their name or mama (e.g., “Which picture starts with the first letter of your name: snake, bear, or duck?”; see Figure 1d). All sessions start with an attractive animation in which preschoolers Sim and Sanne explain the upcoming game; for instance, Sim and Sanne discover that their names start with the same sound.

Tutor. In the tutor condition (LL-Tutor), children received increasingly supportive oral feedback from the tutor to their responses. Unlike most computer games, the program Living Letters gives adultlike feedback that goes beyond indicating which responses are correct and which ones are not. The computer tutor explains why a response is correct (e.g., “Listen, in snake you can hear the /s/ of Sam”). Furthermore, help from the tutor includes more clues as more errors are made in an assignment. For example, (1) after the first error in an assignment, the oral instruction is repeated and children are encouraged to “listen carefully” to promote more thoughtful responses; (2) after the second error, the program provides oral cues to solve the task correctly (e.g., “Do you remember how the teacher writes your name?”), thus enabling solution of the task and engagement in other, similar tasks independently; and (3) a third error is followed by the correct solution with an oral explanation (e.g., “Listen, in that word you can hear the /p/ of Peter”). All tutoring was provided by Sim’s teddy bear (the tutor), as can be seen in Figure 1b.

In the nontutoring condition (LL-NoTutor) children were exposed to Living Letters without an online tutor. Instructions and assignments, as well as the number of repetitions, were the same as
Table 1
Descriptives of Treatment (Living Letters With and Without a Tutor) and Control Groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>LL-NoTutor</th>
<th>Control</th>
<th>LL-Tutor</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Range</td>
<td>N</td>
</tr>
<tr>
<td>Gender (1 = female)</td>
<td>.37</td>
<td>.49</td>
<td>0 to 1</td>
<td>78</td>
</tr>
<tr>
<td>Age in months (fall)</td>
<td>52.40</td>
<td>3.25</td>
<td>48 to 59</td>
<td>78</td>
</tr>
<tr>
<td>Maternal education (highest)</td>
<td>3.16</td>
<td>1.31</td>
<td>1 to 6</td>
<td>64</td>
</tr>
<tr>
<td>SON mosaic (raw scores, winter)</td>
<td>8.16</td>
<td>2.00</td>
<td>3.5 to 15.0</td>
<td>74</td>
</tr>
<tr>
<td>PPVT (raw scores, winter)</td>
<td>67.11</td>
<td>12.26</td>
<td>37 to 104</td>
<td>75</td>
</tr>
<tr>
<td>PPVT (norm scores, winter)</td>
<td>101.96</td>
<td>15.38</td>
<td>64 to 145</td>
<td>75</td>
</tr>
</tbody>
</table>

Note. Screening was done in the fall, pretest in the winter, and posttest in the spring. LL-NoTutor = Living Letters without a tutor; LL-Tutor = Living Letters with a tutor; SON = Snijders-Oomen Niet-verbale intelligentie toets (Snijders-Oomen Nonverbal Intelligence Test); PPVT = Peabody Picture Vocabulary Test.

* Principle components analysis applied to the Stroop-like and digit span tasks revealed two components: working memory (high loadings of digit span tasks and of working memory errors in dogs) and inhibitory control (high loadings of opposites and of inhibitory control errors in dogs).

in the condition with a tutor. Similar to the LL-Tutor condition, an assignment was repeated once or twice after one or two errors; however, there were no comments from a computer tutor. In the LL-NoTutor condition, the game was repeated after an error. In this way, children could be aware of their correct responses and errors.

After a maximum of three trials, Sim, Sanne, and the teddy bear started dancing to mark the end of an assignment, whether or not the child had made an error in an assignment, the game was repeated not only in the same session but also in the next session, with a maximum of two repetitions per game. Thus, children received a variable number of sessions. The total number of sessions, each including six games, ranged from seven to 17, with a mean number of 11.2 sessions (SD = 1.88), each lasting about 10 min. Children in the LL-Tutor and LL-NoTutor conditions participated in an equal number of sessions, \(t(150) = 1.1, p = .29\).

Clever Together. The control group played with another Web-based program, called Clever Together (www.Samenslim.nl). In this program, Sim and Sanne, the same characters as in Living Letters, play hide and seek games. In 40 games of different difficulty levels, children had to help Sim by finding Sanne behind objects. For instance, children were told by the computer voice that Sanne would hide behind a red object. The total number of 10-min sessions ranged from seven to 13. The mean number of 8.01 sessions (SD = 1.01) was somewhat lower than the 11.2 sessions in the intervention group, \(t(299) = 17.4, p < .01\).
students who were blind to group allocation. The order of the tests was always the same, except for the testing of regulatory skills, which were tested in counterbalanced order. Regulatory skills sessions were videotaped and scored afterward by master’s-level students who were blind to group allocation.

**Measures**

**Maternal education.** Mothers reported their highest level of education on a 6-point scale ranging from primary education to university.

**Intelligence.** Verbal intelligence was tested with the Dutch version of the Peabody Picture Vocabulary Test (PPVT; Schlichting, 2005), and nonverbal intelligence was tested with the subtest mosaic of the standardized Snijders-Oomen Niet-verbale Intelligente toets (SON; Snijders-Oomen Nonverbal Intelligence Test; Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998).

**Code-related skills.**

**Developmental spelling.** Children had to write five dictated words (i.e., papa [daddy], Sim [the name of one of the characters in the computer games], been [leg], jurk [dress], and a word starting with the first letter of the child’s name or mama) that afterward were assigned one of the following codes (Levin & Bus, 2003): 0 (drawing-like scribble), 1 (writing-like scribbles, but not similar to conventional symbols), 2 (conventional symbols not representing sounds in the word), 3 (one phonetic letter), 4 (two or more phonetic letters), 5 (invented spelling [readable but not spelled correctly]), and 6 (conventional spelling). All words were double-coded with high Kappa’s (ranging from .88 to .97). Disagreements were solved by discussion. For pre- and posttest, scores on five words were averaged, resulting in a 0–6 scale (as > .84).

**Name–letter knowledge.** Children had to first point to the first letter of their name among five other letters. With few exceptions, all children were able to complete this task successfully. Children then had to name or provide the sound for the first letter of their name. One point was awarded for naming or sounding the correct letter.

**Phonemic sensitivity.** In the phonemic sensitivity task, children were asked to point to the picture of a word that started with or contained the same sound as their name (or mama, for children with an irregular first name letter). The computer named the three optional pictures. A total score of 6 was possible, one for each correct item ($\alpha = .62$).

**Aggregate measure.** Principal components analysis of developmental spelling, name–letter knowledge, and phonemic sensitivity revealed one component for pretests and posttests explaining

![Figure 1](image.png) These screenshots have been derived from four different games: selecting the name (a), selecting *mama* (b), selecting the first letter of the name (c), and selecting the painting that starts with the letter of the child’s own first name (e.g., Sam—snake) among three alternatives (d). When the computer mouse skims a picture, as in d, the computer pronounces the picture’s name.
53% and 55% of the variance, respectively, with high loadings ranging from .62 for name–letter knowledge to .80 for developmental spelling. The aggregate measure for pretest scores was used as covariate, and for posttest scores were used as dependent variable.

**Regulatory skills.**

**Knock and tap.** Regulatory skills at screening were measured with the knock and tap task, in which children had to knock on the table when the experimenter tapped, and vice versa (e.g., Klenberg, Korkman, & Lahtti-Nuuttila, 2001). Similar to the head-to-toes task (Cameron Ponitz et al., 2008), this task is an easy-to-administer measure of behavioral regulation that can be used with very young children. It requires children to pay attention, use their working memory, and inhibit a natural tendency to mimic the experimenter. The internal consistency of this 16-item test was high ($\alpha = .92$).

**Stroop-like task (opposites).** Children had to respond with the opposite to three contrasting pairs of pictures (e.g., saying “fat” to thin) in a mixed set of 18 pictures (based on Berlin & Bohlin, 2002). Incorrect naming and corrections were both scored as errors in this inhibitory control test, with a maximum score of 18 ($\alpha = .91$).

**Stroop-like task (dogs).** Following the Stroop paradigm, children had to switch rules by responding with an opposite (i.e., saying “blue” to a red dog and “red” to a blue dog; based on Beveridge, Jarrold, & Pettit, 2002). The task consisted of 96 trials distributed over four conditions, in which demands on working memory (remembering the name of one or two dogs) and inhibitory control of the most obvious response varied. In the first two conditions, children had to name one or, respectively, two dogs (“Tim” and “Jet”) different in color (yellow and green). In the third and fourth condition, the paradigm was the same; however, the colors of the dogs were incompatible with their names (a red dog was called “blue,” and a blue dog “red”). Incorrect naming and no response were considered as working memory errors, while corrections were scored as inhibitory control errors. Each error was coded as working memory or inhibitory control error, resulting in maximum scores of 96 for both. Internal consistencies for scales were high ($\alpha = .80$ to .94, respectively).

**Digit span (words).** In the forward digit span test (Leidse Diagnosticste Test; Schroots & Van Alphen de Veer, 1976), the children had to repeat a list of unrelated words that was read aloud by the computer. Practice trials were two-word lists. In the test trials, the number of words per list increased from two to a maximum of five and ended when children failed to succeed three series in succession. The total number of correct responses (maximum 12) was the score for this verbal working memory task.

**Backward digit span.** In the backward digit span test (Wechsler Intelligentscale für Kinder—Third Edition; Wechsler, 1992), children had to repeat a string of digits in reverse order. During four practice trials with strings of two to four digits, the experimenter corrected children when needed. The test started with two digits and gradually increased in the number of digits. In each trial, there were two strings of digits, and at least one of these strings had to be repeated correctly in order to proceed to the next trial. The total score for this working memory task was composed of the total number of correct responses in the practice and test trials (maximum 14). Intraclass correlation coefficients between two independent coders were high for all tasks ($rs > .97$).

**Aggregate measures.** Principle components analysis applied to the Stroop-like and digit span tasks revealed two components for regulatory skills in the spring, with high loadings (.63–.86) explaining 34% and 28% of the variance, respectively. The two components can be labeled as working memory (high loadings of digit span tasks and of working memory errors in dogs) and inhibitory control (high loadings of opposites and of inhibitory control errors in dogs).

**Number of trials.** Based on automatic computer registration and storage of mouse behavior during each session, the number of trials each child needed to give a correct answer for within the games was determined. More trials indicated more errors in completing the computer tasks.

**Data Analyses**

Because participants were recruited from different schools ($N = 15$), we used Huber-White estimates of standard errors to correct for clustering of the scores of children from the same schools (cf. Hatcher et al., 2006; Knafo, Israel, & Ebstein, 2011). We included the corrected standard errors in the Complex Sample General Linear Model (SPSS 17), with the posttest score on the aggregate measure (a compound of code-related skills) as dependent variable; experimental condition (LL-NoTutor, control, LL-Tutor) as factor; and age, pretest compound of code-related skills, maternal education, PPVT, SON, inhibitory control, and working memory as covariates. We further examined interactions between experimental condition and regulatory skills.

**Results**

**Attrition**

Nine children moved during the school year. In the remaining group ($N = 303$), one child assigned to the LL-Tutor condition refused to play the games of Living Letters after three sessions. This child was excluded from the final analyses.

**Intervention Effects**

Table 1 presents descriptive statistics for intervention and control conditions on all measures. The three groups were similar in age, maternal education, regulatory skills, and verbal (PPVT; Schlichting, 2005) and nonverbal (SON; Tellegen et al., 1998) intelligence. Correlations between predictors were low to moderate, as is shown in Table 2.

The regression model explained 61% of the variance of code-related skills (see Table 3). The pretest score ($\beta = .62$, 95% confidence interval [CI] [.48, .76]), $t(14) = 9.43$, $p < .001$, was a significant covariate, while working memory ($\beta = .17$, 95% CI [−.00, .35]), $t(14) = 1.94$, $p = .07$, were marginally significant. The background variables (age, maternal educational level, PPVT, and SON) were nonsignificant covariates ($rs = .29$ to .81). Planned contrasts between experimental conditions revealed effects for control group versus LL-Tutor ($\beta = -.38$, 95% CI [−.59, −.16]), $t(14) = 3.75$, $p = .002$, and for LL-NoTutor versus LL-Tutor ($\beta = -.48$, 95% CI [−.66, −.30]), $t(14) = 5.58$, $p < .001$, but not for control group versus LL-NoTutor ($\beta = .10$, .91).
95% CI [–0.10, 0.31]), t(14) = 1.10, p = .29. After using the Šidák-Bonferroni correction (α = .017) to control for Type 1 error rate (Keppel & Wickens, 2004), both contrasts with LL-Tutor remained significant. Contrast the target programs with the control condition, effect sizes were d = 0.48 with tutor and d = −0.14 without. The difference between LL-NoTutor and LL-Tutor equaled .71 standard deviation (see Table 3).

Did regulatory skills moderate intervention effects? We found a significant interaction between the contrast LL-NoTutor versus LL-Tutor and inhibitory control (β = .17, 95% CI [0.02, .31]), t(14) = 2.49, p = .03. This indicates that with an online tutor the achievement gap between children with low and high inhibitory control was small (see Figure 2), while without online tutoring the gap between children with low and high inhibitory control increased. According to separate regression analyses, effect sizes of inhibitory control were d = 0.49 (p = .07) within the LL-Tutor condition and d = 0.67 (p = .02) within the LL-NoTutor condition.

### Number of Trials

There were differences in the average number of trials that children needed to solve the games (later referred to as “trials”). Children in the LL-Tutor condition needed fewer trials preceding a correct answer than did children in the LL-NoTutor condition, t(158) = 2.0, p = .045, which may indicate that the presence of a tutor discouraged random response behavior and thus errors. In the LL-Tutor and the LL-NoTutor conditions, children needed 1.78 (SD = 0.40) and 1.92 (SD = 0.47) trials, respectively. Table 2 shows that number of trials was negatively correlated with work-

---

### Table 2

**Correlations Between All Included Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gendera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>−.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal education</td>
<td>−.11</td>
<td>−.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPVT</td>
<td>−.11</td>
<td>.30**</td>
<td>.27**</td>
<td>.39**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre code-related skillsb</td>
<td>.15**</td>
<td>.36**</td>
<td>.12</td>
<td>.36**</td>
<td>.45**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post code-related skillsb</td>
<td>.11</td>
<td>.26**</td>
<td>.15*</td>
<td>.34**</td>
<td>.40**</td>
<td>.72**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knock and tap</td>
<td>.07</td>
<td>.10</td>
<td>.06</td>
<td>.16*</td>
<td>.20**</td>
<td>.19*</td>
<td>.13*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working memoryb</td>
<td>−.03</td>
<td>.25**</td>
<td>.16**</td>
<td>.38**</td>
<td>.43**</td>
<td>.50**</td>
<td>.46**</td>
<td>.19**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inhibitory controlb</td>
<td>.05</td>
<td>.20**</td>
<td>.07</td>
<td>.27**</td>
<td>.29**</td>
<td>.25**</td>
<td>.28**</td>
<td>.18**</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trials (computer game)</td>
<td>−.05</td>
<td>−.08</td>
<td>−.10</td>
<td>−14*</td>
<td>−.06</td>
<td>−.18**</td>
<td>−.22*</td>
<td>−.07</td>
<td>−14*</td>
<td>−.22**</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* N varies between 254 and 312. SON = Snijders-Oomen Niet-verbale Intelligentie toets (Snijders-Oomen Nonverbal Intelligence Test; raw scores of the subtest mosaic); PPVT = Peabody Picture Vocabulary Test (raw scores).

a Gender (0 = boy, 1 = girl).  
* Aggregate measures.

p < .05. ** p < .01.

---

### Table 3

**Results of Regression Analysis With Posttest Code-Related Skills as Dependent Measure**

<table>
<thead>
<tr>
<th>Measure</th>
<th>β (SE)</th>
<th>95% CI</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>−.02 (.01)</td>
<td>[−.05, 0.02]</td>
<td>−1.10</td>
<td>.29</td>
<td>−.14</td>
</tr>
<tr>
<td>Maternal education</td>
<td>.01 (.03)</td>
<td>[−.06, 0.08]</td>
<td>0.25</td>
<td>.81</td>
<td>.03</td>
</tr>
<tr>
<td>PPVT</td>
<td>.00 (.00)</td>
<td>[−.01, 0.01]</td>
<td>0.61</td>
<td>.55</td>
<td>.08</td>
</tr>
<tr>
<td>SON</td>
<td>.01 (.03)</td>
<td>[−.05, 0.07]</td>
<td>0.36</td>
<td>.73</td>
<td>.05</td>
</tr>
<tr>
<td>Pretest code-related skills</td>
<td>.62 (.07)</td>
<td>[0.48, 0.76]</td>
<td>9.43</td>
<td>.00</td>
<td>1.20</td>
</tr>
<tr>
<td>Main effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working memory</td>
<td>.17 (.08)</td>
<td>[−.00, 0.35]</td>
<td>2.10</td>
<td>.06</td>
<td>.27</td>
</tr>
<tr>
<td>Inhibitory control</td>
<td>.13 (.07)</td>
<td>[−.01, 0.27]</td>
<td>1.96</td>
<td>.07</td>
<td>.25</td>
</tr>
<tr>
<td>C1</td>
<td>−.48 (.09)</td>
<td>[−.66, −.30]</td>
<td>−5.58</td>
<td>.00</td>
<td>−.71</td>
</tr>
<tr>
<td>C2</td>
<td>−.38 (.10)</td>
<td>[−.59, −.16]</td>
<td>−3.75</td>
<td>.00</td>
<td>−.48</td>
</tr>
<tr>
<td>C3a</td>
<td>.10 (.09)</td>
<td>[−.10, 0.31]</td>
<td>1.10</td>
<td>.29</td>
<td>.14</td>
</tr>
<tr>
<td>Interaction effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1 × Inhibitory Control</td>
<td>.17 (.07)</td>
<td>[0.02, 0.31]</td>
<td>2.49</td>
<td>.03</td>
<td>.32</td>
</tr>
<tr>
<td>C2 × Inhibitory Control</td>
<td>−.04 (.08)</td>
<td>[−.20, 0.13]</td>
<td>−0.45</td>
<td>.66</td>
<td>−.06</td>
</tr>
<tr>
<td>C1 × Working Memory</td>
<td>.03 (.12)</td>
<td>[−.23, 0.29]</td>
<td>0.27</td>
<td>.79</td>
<td>.03</td>
</tr>
<tr>
<td>C2 × Working Memory</td>
<td>−.09 (.11)</td>
<td>[−.32, −.14]</td>
<td>−0.82</td>
<td>.42</td>
<td>−.10</td>
</tr>
</tbody>
</table>

*Note.* N = 248. For calculating Cohen’s d we used the formula 2rt/n − 2 (Thalheimer & Cook, 2002). CI = confidence interval; PPVT = Peabody Picture Vocabulary Test; SON = Snijders-Oomen Niet-verbale Intelligentie toets (Snijders-Oomen Nonverbal Intelligence Test); C1 = contrast between Living Letters without a tutor and Living Letters with a tutor; C2 = contrast between control group and Living Letters with a tutor; C3 = contrast between control group and Living Letters without a tutor.

a Effect was calculated in a separate analysis.
ing memory ($r = -.14$) and inhibitory control ($r = -.22$), which may indicate that children with low regulatory skills were more inclined to respond randomly, thereby making more errors.

**Discussion**

An intelligent tutoring system, modeled after early literacy activities in literate homes (Living Letters with tutor), was shown to improve literacy skills from children of low-educated families. Four-year-olds’ code-related skills improved substantially when they were exposed to this computer program with minimal supervision by teachers. In this study, more importantly, games were found to be effective only when an online tutor was used to explain how to proceed and why the solution was correct. Consistent with prior research (Azevedo & Bernard, 1995; Meyer et al., 2010; Vasilyeva, 2007), the effects of instructions and assignments were reduced when children did not receive immediate and personalized reactions to their game responses. Children in the tutoring condition outperformed the children in the nontutoring condition by about three quarters of a standard deviation, supporting the hypothesis that computer programs for young children need built-in tutors (Bodrova & Leong, 2006). It was an unexpected result that the intervention group that did not receive immediate and personalized reactions from an online tutor scored lower than did the control group. Even though the difference did not reach statistical significance, this result may indicate that Living Letters without an online tutor can have negative effects on problem solving. We hypothesized that a computer program that does not provide tutoring modeled on adult caregivers may reward random responses instead of strengthening thoughtful replies. Without online tutoring, children may be more inclined to guess when they have to solve similar problems during posttest assessments. The on-average higher number of trials to select correct answers in the condition without a tutor corroborates this hypothesis. With online tutors young children may take the computer assignments more seriously and prefer a reflective approach to random clicking and answering.

Our findings are consistent with earlier work showing that regulatory skills are predictors of academic achievements (e.g., Bierman et al., 2008; Blair & Razza, 2007; Davidse et al., 2011; Kegel et al., 2009; McClelland et al., 2007). However, the current results also nuance the importance of regulatory skills. There was no evidence for effects of working memory on code-related skills. Only inhibitory control affected gains in code-related skills. Furthermore, there was no evidence that inhibitory control affects learning across computer environments to the same extent. Inhibitory control had a marginally significant effect on gains in code-related skills in the ITS game where a tutor corrected or confirmed children’s responses after each game. In this tutoring condition,
inhibitory control explained 6% of the posttest differences. In the nontutoring condition, however, the group high on inhibitory control outperformed the group low on inhibitory control. In this condition, inhibitory control explained 10% of the posttest scores. In fact, the outcomes thus evidence a dual-risk model (Belsky et al., 2007): Children with some risk (here: low inhibitory control) lagged further behind when they were exposed to a less-supportive environment (here: NoTutor), while children with low and high inhibitory control benefited to the same extent from a supportive environment (here: Tutor). The best explanation for the effect of inhibitory control in the condition without a tutor seems to be trial and error behavior. When children are easily distracted by irrelevant cues, they may finalize the assignments by clicking randomly without any reflection on questions, which matches our finding that they tend to need more trials preceding the correct solution. It should be noted, however, that differences between children with high and low inhibitory control were rather small, probably because average progress in the condition without an online tutor was minor.

Limitations

This study has some limitations. We used the knock and tap task to compose three experimental groups similar in regulatory skills. Attempts to apply more-complicated tests of regulatory skills before the intervention failed because they appeared to be too demanding for young children (e.g., dimensional change card sort task; Zelazo, 2006). Yet, to test which regulatory skills might affect gains and moderate the effects of the program, another more-extensive set of measures of regulatory skills was applied after the intervention and used in the final analyses to test main and interaction effects of regulatory skills. Scores on the knock and tap task as well as on other measures of regulatory skills were similar across experimental groups. However, some might argue that the intervention may have changed regulatory skills and that assessment after the intervention is less appropriate to test the effects of regulatory skills on the intervention.

Another limitation is that the current study provides only short-term evidence. Long-term evidence is needed to demonstrate that computer programs targeting early literacy and including an online tutor can be a preemptive measure in preschool. Also, because of the moderate effect sizes of the intervention, one might expect that there may be other, not yet considered individual differences that make one group of children more susceptible to interventions than another, a hypothesis to be tested in further research (Belsky et al., 2007).

Implications

Computer instruction seems to be a promising addition to classroom instruction in particular when the programs include an online tutor that corrects children’s responses and provides cues. Traditional measures of school readiness focus primarily on preacademic skills, such as emergent reading and writing, and less on behavioral skills (e.g., Duncan et al., 2007). The present findings are consistent with the earlier work showing that children’s regulatory skills are important in addition to cognitive measures (Kegel et al., 2009). The current results evidence that especially children with underdeveloped inhibitory control score disproportionately negatively in a less-supportive computer environment. Although the reported effects of inhibitory control on learning were rather small, findings make plausible the idea that inhibitory control is an important explanation for outcomes of learning via computers, especially when one considers that a lot of current computer programs lack tutoring.

References

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool


Received December 21, 2010
Revision received August 2, 2011
Accepted September 13, 2011