What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

Beginning Reading September 28, 2006

DaisyQuest

Program description

DaisyQuest is a software bundle that offers computer-assisted instruction in phonological awareness, targeting children aged three to seven years. The instructional activities, framed in a fairy tale involving a search for a friendly dragon named Daisy, teach children how to recognize words that rhyme; words that have the same beginning, middle, and ending sounds; and words that can be formed from a series of phonemes presented separately, as well as how to count the number of sounds in words.

Research

Four studies met the What Works Clearinghouse (WWC) evidence standards. The studies included a total of 223 students ranging in age from five to seven years, attending schools in different communities and states, including one western and one southeastern state. The studies examined *DaisyQuest's* effects in the alphabetics domain, specifically on phonological awareness and phonics measures.[1]

Effectiveness

DaisyQuest was found to have positive effects on alphabetics skills.

	Alphabetics	*Fluency*	*Comprehension*	*General reading achievement*
Rating of effectiveness	Positive	Not reported	Not reported	Not reported
Improvement index[2]	Average: +23 percentile points Range: –18 to +45 percentile points	Not reported	Not reported	Not reported

1. The evidence presented in this report is based on the available research. Findings and conclusions may change as new research becomes available.
2. The improvement index is based on the average effect size within a domain, and the range of improvement indices represents the minimum and maximum of all individual findings across all studies.

Additional program information

Developer and contact

Gina C. Erickson (DaisyQuest@comcast.net)

Scope of use

DaisyQuest was developed in 1992. Information is not available on the number or demographics of students, schools, or districts using the software.

Teaching

DaisyQuest is a software program comprised of two components that focus on recognition of rhyming; recognizing words with the same beginning, middle, and ending sounds; forming words from a series of phonemes; and counting sounds in words. The software uses graphics and story lines to engage children in the learning process. In the first component, called *DaisyQuest*, as children master each level of instructional activities they are rewarded with clues that lead them to discover where Daisy is hiding. In the second component, *Daisy's Castle,* a similar fairy tale theme involves searching for Daisy's lost eggs. The programs also offer children choices about the sequence of instructional activities and keep track of children's responses. Materials are presented using digitized and synthetic speech. The software contains a tutorial that guides the child by explaining each skill or concept briefly and provides practice exercises with feedback for correct and incorrect responses. When the activity is completed the child's mastery of the concept is tested through activities and questions. Included with the program is an adaptive test called *Undersea Challenge.* This test measures children's knowledge of rhyming; beginning, middle, and ending sounds; and phoneme blending and segmenting. The software generates statistical reports that enable parents and teachers to view children's performance.

In each of the four studies reviewed, classroom teachers did not direct students' use of *DaisyQuest* (studies were conducted by experimenters). However, the software is self-contained and teachers may send students to the computer to practice these skills, without any need to implement additional curricular materials. Though not stipulated by the program developer, the students in the reviewed studies used the computer program for 15–32 sessions, each lasting 20–25 minutes.

Cost

The *DaisyQuest* bundle (*DaisyQuest, Daisy's Castle,* and the *Undersea Challenge* mastery test) is available for $49.95, plus $6.95 shipping and handling.

Research

Five studies reviewed by the WWC investigated the effects of the *DaisyQuest* program. Four studies (Barker & Torgesen, 1995; Foster, Erickson, Foster, Brinkman, & Torgesen, 1994, Experiment 1: Child-care Facility; Foster, Erickson, Foster, Brinkman, & Torgesen, 1994, Experiment 2: Kindergarten Classrooms; and Mitchell & Fox, 2001) were randomized controlled trials that met WWC evidence standards. The one remaining study did not meet WWC evidence screens. In two of the studies, the authors used two groups to make comparisons to the intervention group. The beginning reading review presents data relevant to all comparisons.[3]

Barker and Torgesen (1995) presented results for 49 at-risk first graders who had been randomly assigned to one of three conditions: *DaisyQuest, Hint and Hunt,* a software program that teaches short vowel sounds, or computer time to work with math-oriented software programs.[4]

Foster et al. (1994, Experiment 1: Child-care Facility) analyzed outcomes of 27 students randomly assigned to use *DaisyQuest* or to continue receiving only their regular school program.

Foster et al. (1994, Experiment 2: Kindergarten Classrooms) examined results for 69 kindergarteners who were randomly

3. Considering all comparison groups, rather than selecting one comparison over another, is especially important in a topic such as reading, where there is generally no true control (absence of reading instruction).

4. In this and the Mitchell and Fox (2001) study, the authors include students working with math and drawing software programs to account for the possibility that working with any software programs may improve reading skills.

Research *(continued)*

assigned to one of two conditions: *DaisyQuest* or their regular kindergarten curriculum.

Mitchell and Fox (2001) focused on 69 students randomly assigned to one of three conditions: *DaisyQuest* (intervention), teacher-delivered phonological awareness instruction, where teachers guided students through oral activities (comparison 1), or mathematics and drawing software programs (comparison 2).

Effectiveness

Findings

The WWC review of beginning reading addresses student outcomes in four domains: alphabetics, reading fluency, comprehension, and general reading achievement.[5] *DaisyQuest* studies addressed outcomes in alphabetics and included outcomes for two constructs within alphabetics—phonological awareness and phonics. All four *DaisyQuest* studies (Barker & Torgesen, 1995; Foster et al., 1994, Experiment 1: Child-care Facility; Foster et al., 1994, Experiment 2: Kindergarten Classrooms; Mitchell & Fox, 2001) used phonological awareness measures. Barker and Torgesen (1995) also used phonics measures. The findings below present authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *DaisyQuest* on students. Sometimes the two differ, reflecting WWC calculations based on data provided by the authors (see Appendix A3).[6]

Alphabetics. The Barker and Torgesen (1995) study findings are based on the performance of *DaisyQuest* students and comparison students on five measures of **phonological awareness** and four measures for **phonics** in each set of comparisons.

When the *DaisyQuest* group was compared with the alternative reading software group, the study authors found statistically significant effects favoring the *DaisyQuest* group for three of the five **phonological awareness** measures. The WWC analysis found that two of five positive effects for **phonological awareness** (*Undersea Challenge* and Production Test of Segmenting) were statistically significant. One additional positive effect (Phoneme Elision Test), while not statistically significant, was large enough to be considered substantively important according to WWC criteria.[7] The study authors also found a statistically significant effect favoring the *DaisyQuest* group for one of the four **phonics** measures (Woodcock-Johnson Word Identification subtest). The WWC effect size computations found none of the four positive effects for **phonics** to be statistically significant; but three effects (Woodcock-Johnson Word Identification subtest, Woodcock-Johnson Word Analysis, and Experimental Non-word Reading) were large enough to be considered substantively important according to WWC criteria.

When the *DaisyQuest* group was compared with the math-oriented software, the study authors found, and the WWC confirmed, statistically significant effects favoring the *DaisyQuest* group for two of the five **phonological awareness** measures (*Undersea Challenge* and Production Test of Segmenting). The other three positive effects (Phoneme Elision Test, Sound Categorization, and Production Test of Blending), while not statistically significant, were large enough to be considered substantively important according to WWC criteria. The study authors also found a statistically significant effect favoring the *DaisyQuest* group for one of the four **phonics** measures (Woodcock-Johnson Word Identification subtest) but this was not confirmed by the WWC. According to WWC calculations, there were three positive effects (Woodcock-Johnson Word

5. For definitions of the domains, see the Beginning Reading Protocol.
6. The level of statistical significance was calculated by the WWC and, where necessary, corrects for clustering within classrooms or schools, for multiple outcomes within one domain, and for multiple comparisons. For an explanation see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of *DaisyQuest,* corrections for multiple outcomes and for multiple comparison groups were needed.
7. A substantively important effect is defined as an effect size greater than positive or negative 0.25.

Effectiveness *(continued)*

Identification subtest, Woodcock-Johnson Word Analysis, and Experimental Non-word Reading) that, while not statistically significant, were large enough to be considered substantively important. The one negative effect (Analog Reading Task) found by the WWC was neither statistically significant nor substantively important according to WWC criteria.

Foster et al. (1994, Experiment 1: Child-care Facility) used two phonological awareness tests (phonological awareness test (b) and the screening test of phonological awareness–experimental version). For both measures, the authors found, and the WWC confirmed, positive, statistically significant effects in favor of the *DaisyQuest* group.

Foster et al. (1994, Experiment 2: Kindergarten Classrooms) used four phonological awareness tests. The authors found, and the WWC confirmed, positive, statistically significant effects favoring the *DaisyQuest* group on three measures (*Undersea Challenge,* Production Test of Segmenting, and Production Test of Blending).

Mitchell and Fox (2001) used the Phonological Awareness Test (a), which included a total test score and four subtests for each of the two comparison groups. Only the total test score was included in the effectiveness rating.[8] In the comparison between the *DaisyQuest* group and the teacher-delivered phonological awareness group, the authors found no statistical differences on the total test score. According to WWC effect size computations, there was a negative effect—that is, the *DaisyQuest* group scored lower than the teacher-led group on this measure. Although the effect was not statistically significant, it was large enough to be substantively important by WWC standards. In the comparison of *DaisyQuest* students and students using the other instructional technology, the study authors found, and the WWC confirmed, statistically significant positive effects on the total test score for the *DaisyQuest* group.

For alphabetics, three studies were categorized as having positive effects and had strong designs. One study had a strong design and was categorized as having indeterminate effects.

Rating of effectiveness

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings (as calculated by the WWC), the size of the difference between participants in the intervention condition and the comparison condition, and the consistency in findings across studies (see the WWC Intervention Rating Scheme). Overall, the WWC found *DaisyQuest* to have postive effects for alphabetics.

The WWC found *DaisyQuest* to have positive effects for alphabetics

Improvement index

For each outcome domain, the WWC computed an improvement index based on the effect size (see the Technical Details of WWC-Conducted Computations). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results. The improvement index range across all individual outcomes for alphabetics is –18 to +45 percentile points. The improvement index based on the domain average effect size across all studies is +23 percentile points. That is, the average student in the comparison group would be expected to improve from the 50th percentile to the 73rd percentile after receiving the intervention.

8. The WWC does not include subtests in the effectiveness ratings of each study to avoid counting one test multiple times. But effect size estimates on the subtests are presented in Appendix A4.1.

The WWC found *DaisyQuest* to have positive effects for alphabetics *(continued)*

Summary

The WWC reviewed four studies on *DaisyQuest* that met WWC evidence standards; each focused on outcomes in the alphabetics domain. The WWC categorized three of the four studies as having statistically significant effects and one study as having indeterminate effects. So, when the WWC looked at results from *DaisyQuest* in aggregate, the intervention was found to have positive effects on alphabetics. The evidence in this report is limited and may change as new research become available.

References

Met WWC evidence standards

Barker, T., & Torgesen, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research, 13*(1), 89–103.

Foster, K. C., Erickson, G. C., Foster, D. F., Brinkman, D., & Torgesen, J. K. (1994). Computer administered instruction in phonological awareness: Evaluation of the *DaisyQuest* program. *Journal of Research and Development in Education, 27*(2), 126–137. (Experiment 1: Child-care Facility).

Foster, K. C., Erickson, G. C., Foster, D. F., Brinkman, D., & Torgesen, J. K. (1994). Computer administered instruction in phonological awareness: Evaluation of the *DaisyQuest* program. *Journal of Research and Development in Education, 27*(2), 126–137. (Experiment 2: Kindergarten Classrooms).

Mitchell, M. J., & Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction, 40*(4), 315–332.

Did not meet WWC evidence standards

Lonigan, C. J., Driscoll, K., Phillips, B. M., Cantor, B. G., Anthony, J. L., & Goldstein, H. (2003). A computer-assisted instruction phonological sensitivity program for preschool children at-risk for reading problems. *Journal of Early Intervention, 25*(4), 248–262.[9]

For more information about specific studies and WWC calculations, please see the WWC *DaisyQuest* Technical Appendices.

9. The age of the students in this study was outside the range of this review.

Appendix

Appendix A1.1 Study characteristics: Barker & Torgesen, 1995 (randomized controlled trial)

Characteristic	Description
Study citation	Barker, T., & Torgesen, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research, 13*(1), 89–103.
Participants	Participants were the 54 students who met eligibility criteria (scoring below the 40th percentile on the Woodcock-Johnson Word Identification subtest) and the sound categorization measure (below 50th percentile). Initially, 87 at-risk first graders (approximately 6–7 years old) were nominated by their teachers and screened for study eligibility. The 54 qualifying students were given additional pretests and then randomly assigned to either the intervention or comparison group. Due to attrition, 49 students were in the final analysis sample.
Setting	This study took place at two elementary schools.
Intervention	Intervention students used the *DaisyQuest* software in a school psychologist's office in groups of three or four students under the direction of an experimenter. Students wore headphones and used the software independently during intervention sessions that lasted 25 minutes four times a week for eight weeks. Students used both components of the *DaisyQuest* program. This version of *DaisyQuest* contained seven instructional activities.
Comparison	Two comparison groups were used. Both sets of students used computers for the same amount of time as the intervention group for either an alphabetic decoding program that focused on vowel sounds (*Hint and Hunt*) or computer-based math programs (including *Alien Addition, Math Rabbit,* and *Math Blaster*).
Primary outcomes and measurement	The authors used a battery of tests for pre- and posttests. The Woodcock-Johnson Reading Mastery Word Identification subtest and a sound categorization measure were used as screening measures for eligibility at pretest and as posttests. Students assigned to the study were given an additional seven tests as pre- and posttests: *Undersea Challenge*, the Woodcock-Johnson Reading Mastery Word Analysis subtest, a phoneme elision task, a production test of segmenting, a production test of blending, experimental nonword reading, and an analog reading task. The vocabulary measure from the Stanford Binet IV-Revised test was also mentioned by authors, but results for this measure were not presented. (See Appendix A2 for a more detailed description of outcome measures.)
Teacher training	No information was given about teacher training, because teachers did not deliver instruction for any of the groups.

Appendix A1.2 Study characteristics: Foster, Erickson, Foster, Brinkman, & Torgesen, 1994, Experiment 1: Child-care Facility (randomized controlled trial)

Characteristic	Description
Study citation	Foster, K. C., Erickson, G. C., Foster, D. F., Brinkman, D., & Torgesen, J. K. (1994). Computer administered instruction in phonological awareness: Evaluation of the *DaisyQuest* program. *Journal of Research and Development in Education, 27*(2), 126–137. (Experiment 1: Child-care Facility).
Participants	Participants were 27 eligible students who were randomly assigned to an experimental group (n=12) and a control group (n=15). Before the study, a pool of more than 100 five-year-old children was given the PPVT-R and PAT (b).[1] Children with PPVT-R standard scores less than 75 and children with PAT (b) scores greater than 20 were excluded from the study. The two groups of eligible students were not significantly different from one another in terms of age or scores on the two measures. The average age of children in the experimental group was five years five months and in the control group, five years three months. Although the children in this study were recruited from a preschool, they met age requirements of this review (average age was five years). No attrition occurred.
Setting	Children attended the Kinderland Center, a child-care facility in Orem, Utah.
Intervention	Intervention students participated in 20 *DaisyQuest* computer sessions of approximately 20–25 minutes each in quiet rooms where computer interaction could take place without interruption. Sessions were designed so that students could finish three levels of the program. If a child mastered all three levels before 20 sessions had occurred, training was discontinued. The version of *DaisyQuest* evaluated in this study contained six instructional activities.
Comparison	The comparison group remained in their regular classroom, receiving their routine preschool instruction.
Primary outcomes and measurement	Subjects in both groups were given the Phonological Awareness Test (PAT) and Screening Test of Phonological Awareness-Experimental Version (STOPA-E) in a random order after all children in the experimental group had concluded training. The posttests were given approximately one month after the pretest (that is, the study lasted approximately one month). (See Appendix A2 for a more detailed description of outcome measures.)
Teacher training	No information was given about teacher training, because teachers did not deliver the intervention.

1. Two of the studies in this review included identically named but distinct measures. To distinguish between the two, we denote them as PAT (a) and PAT (b). Please see Appendix A2.

Appendix A1.3 Study characteristics: Foster, Erickson, Foster, Brinkman, & Torgesen, 1994, Experiment 2: Kindergarten Classrooms (randomized controlled trial)

Characteristic	Description
Study citation	Foster, K. C., Erickson, G. C., Foster, D. F., Brinkman, D., & Torgesen, J. K. (1994). Computer administered instruction in phonological awareness: Evaluation of the *DaisyQuest* program. *Journal of Research and Development in Education, 27*(2), 126–137. (Experiment 2: Kindergarten Classrooms).
Participants	Participants were 70 second-semester kindergarten students aged five to seven years (average six years old) from four classrooms. Originally, nearly 97 students (all the students) from four kindergarten classrooms in a suburban elementary school were tested with the PPVT-R. Children with the highest and lowest scores were removed to reduce heterogeneity of the sample with regard to verbal ability. The 70 remaining children were matched in pairs according to their scores on the PPVT-R, with one of each pair being randomly assigned to either the experimental group or the control group. One child from the experimental group changed schools and did not complete the study, for an analysis sample of 69.
Setting	The study took place in a suburban elementary school.
Intervention	Intervention students received 16 daily, 20-minute *DaisyQuest* verbal training sessions in groups of four. Sessions took place at computers located in the hallway outside the child's classroom under the guidance of an experimenter, who assisted students with their headphones and any computer glitches. A few of the children were absent from several sessions, hence training time varied from 4.0 to 5.3 hours, with an average of 4.9 hours of training. This version of *DaisyQuest* contained seven instructional activities. Students varied in the number of activities completed and speed with which they finished the activities.
Comparison	The control group remained in their regular classroom, receiving their routine kindergarten instruction.
Primary outcomes and measurement	A series of tests were given at both pre- and posttesting: Screening Test of Phonological Awareness (STOPA), *Undersea Challenge*, Production Test of Blending, and Production Test of Segmenting. (See Appendix A2 for a more detailed description of outcome measures.)
Teacher training	No information was given about teacher training, because teachers did not deliver the intervention.

Appendix A1.4 Study characteristics: Mitchell & Fox, 2001 (randomized controlled trial)

Characteristic	Description
Study citation	Mitchell, M. J., & Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction, 40*(4), 315–332.
Participants	Participants were 72 students (36 kindergarteners and 36 first graders). To determine eligibility, the district-administered Literacy Initiative for Everyone (LIFE, 1996) inventory was used. Kindergarteners who did not meet district criteria on three of the five kindergarten LIFE subtests and first graders who were below grade level expectations on five of the seven first-grade LIFE subtests were then given the PPVT-III. Seventy-two randomly selected students who met the LIFE requirement and received a standard score of 85 or higher on the PPVT-III formed the sample eligible for this study. These students were then randomly assigned to one of three conditions: *DaisyQuest* (intervention), teacher-administered phonological awareness training (comparison 1); or math and drawing software programs (comparison 2). Twenty-four students were assigned to each study group, half kindergarteners and half first graders. Three students total were lost to attrition, for an analysis sample of 69.
Setting	Six kindergarten and six first-grade classrooms in a middle-class, suburban elementary school in a southeastern state.
Intervention	Intervention students used the *DaisyQuest* software over a four-week period, involving 15, 20-minute sessions (five hours instruction total). Each child was assigned a specific computer in the school's computer lab to use for the length of the study and was guided by an experimenter, who helped them with their earphones and any computer gliches. Students used both components of the *DaisyQuest* software.
Comparison	In comparison 1, students also had 15, 20-minute sessions over a four-week period during which teachers guided them through oral activities focusing on rhyming, articulating single syllable words, identifying sounds in isolation, and matching phonemes. Instructional materials for this condition were selected from the Phonological Awareness Kit (Robertson & Salter, as cited in Mitchell & Fox, 2001) and the Phonological Awareness Intermediate Kit (Robertson & Salter, as cited in Mitchell & Fox, 2001). In comparison 2, students interacted with computers for the same time and duration as the intervention group. Instead of using *DaisyQuest,* participants used one drawing program, *Kid Works 2* (Davidson, as cited in Mitchell & Fox, 2001), and four math software programs, *Math Rabbit* (The Learning Company, as cited in Mitchell & Fox, 2001), *Troggle Trouble Math* (MECC, as cited in Mitchell & Fox, 2001), *Number Maze* (Great Wave Software, as cited in Mitchell & Fox, 2001), and *New Math Blasters Plus* (Davidson, as cited in Mitchell & Fox, 2001). Like the intervention group, they were guided by an experimenter while using these programs in a computer lab.
Primary outcomes and measurement	The Phonological Awareness Test (PAT) (a) was administered pre- and posttest. Overall PAT (a) scores, as well as scores on its Rhyming, Isolation, Segmentation, and Blending subtests were reported. (See Appendix A2 for a more detailed description of outcome measures.)
Teacher training	Teachers did not deliver the intervention or comparison 2, so no information was provided. For comparison 1, the study reported that teachers followed procedures from the two kits (see above).

Appendix A2 Outcome measures in the alphabetics domain

Outcome measure	Description
Phonological awareness	
Phoneme Elision Task	The task measures the child's ability to manipulate root words (in compound words), syllables, and phonemes in words. The tester says aloud a word or nonword and asks the child to repeat it. Then the child is asked to say the same word or nonword, omitting a particular root word, syllable, or phoneme (as cited in Barker & Torgesen, 1995).
Phonological Awareness Test (PAT) (a)	The PAT (a) authored by Robertson and Salter (as cited in Mitchell & Fox, 2001) is designed to measure a child's phonological processing. The study authors provided a total score from four subtests of the test: Rhyme Discrimination and Production; Phoneme Isolation; Phoneme Segmentation; and Blending (as cited in Mitchell & Fox, 2001).
Phonological Awareness Test (PAT) (b)	A 30-item test constructed by authors of the study. The test is administered individually in an oral format and assesses children's ability to recognize rhyming words; recognize whether a given word can be formed from a sequence of separately pronounced phoneme; recognize whether two words have the same beginning, middle, and ending sounds; and recognize whether a word contains a given number of different sounds. Children are tested on four to six items for each skill and respond in a yes/no format (as cited in Foster et al., 1994, Experiment 1: Child-care Facility).
Phonological Awareness Test (PAT) (a): Blending subtest	The Blending subtest consists of two tasks—blending of syllables and blending of phonemes. These are used to assess the student's ability to blend units of sound together to form words (as cited in Mitchell & Fox, 2001).
Phonological Awareness Test (PAT) (a): Phoneme Isolation subtest	The Isolation subtest measures a child's ability to identify individual phonemes by isolating phonemes located at the beginning, middle, and end of words (as cited in Mitchell & Fox, 2001).
Phonological Awareness Test (PAT) (a): Phoneme Segmentation subtest	The Segmentation subtest consists of three tasks: sentences, syllables, and phonemes. These tasks measure a child's ability to divide sentences into words, words into syllables, and words into phonemes or sounds (as cited in Mitchell & Fox, 2001).
Phonological Awareness Test (PAT) (a): Rhyme Discrimination and Production subtest	The Rhyming subtest consists of two tasks: discrimination and production. Discrimination measures the child's ability to identify rhyming words presented in pairs. Production measures the child's ability to provide a word that rhymes with a given stimulus word (as cited in Mitchell & Fox, 2001).
Production Test of Blending	The Production Test of Blending is an individually administered task that requires the child to listen to sequences of phonemes presented separately and to pronounce the word that is made when the sounds are blended together. There are 15 words in this task, ranging from two to six phonemes in length (as cited in Barker & Torgesen, 1995 and Foster et al., 1994, Experiment 2: Kindergarten Classrooms).
Production Test of Segmenting	The Production Test of Segmenting is an individually administered task that requires the child to pronounce, in sequence, each of the separate sounds in a word. The tester presents 15 words that are two to five phonemes in length, and the child must explicitly segment the words (as cited in Barker & Torgesen, 1995 and Foster et al., 1994, Experiment 2: Kindergarten Classrooms).
Screening Test of Phonological Awareness (STOPA)	This is the final version of the paper-and-pencil test developed by Torgesen and Bryant (as cited in Foster et al., 1994). This test contains 30 multiple-choice items that require the child to either identify which of three pictured words begins with the same first sound as another pictured word, identify which of four pictured words begins with a different first sound from the others, or count the phonemes in words that are one to three phonemes in length (as cited in Foster et al, 1994, Experiment 2: Kindergarten Classrooms).

(continued)

Appendix A2 Outcome measures in the alphabetics domain *(continued)*

Outcome measure	Description
Phonological awareness *(continued)*	
Screening Test of Phonological Awareness-Experimental Version (STOPA-E)	This is the first version of the paper-and-pencil test developed by Torgesen and Bryant (as cited in Foster et al., 1994). The measure contains 30 multiple-choice items that require the child to either identify which of three pictured words begins with the same first sound as another pictured word, identify which of four pictured words begins with a different first sound from the others, or count the phonemes in words that are one to three phonemes in length (as cited in Foster et al, 1994, Experiment 1: Child-care Facility).
Sound categorization	This task presents the child with arrays of four words and requires that the child select which word contains a different beginning, middle, or ending sound (as cited in Barker & Torgesen, 1995).
Undersea Challenge	*Undersea Challenge* is a computerized-adaptive test, created by the *DaisyQuest* developer, that selects items to present to the test taker based on the child's previous responses. The test measures children's knowledge of rhyming; beginning, middle, and ending sounds; and phoneme blending and segmenting. All seven types of item formats use the same yes/no response scale (as cited in Barker & Torgesen, 1995 and Foster et al., 1994, Experiment 2: Kindergarten Classrooms).
Phonics	
Analog Reading Task	This task was adapted from Byrne and Fielding-Barnsley (as cited in Barker & Torgesen, 1995). The tester presents the child with two printed words and asks which word is the one being pronounced by the tester. Words, printed in lower-case letters on individual index cards, are presented in pairs (as cited in Barker & Torgesen, 1995).
Experimental Non-Word Reading	Similar to the Word Analysis subtest of the Woodcock-Johnson, children are asked to read aloud 15 two- and three-letter nonwords (as cited in Barker & Torgesen, 1995).
Woodcock-Johnson Reading Mastery: Word Analysis subtest	The Word Analysis subtest measures the ability to apply phonics skills to pronounce unfamiliar words. The child is asked to read aloud either nonsense words or words with a very low frequency of occurrence in English (as cited by Barker & Torgesen, 1995).
Woodcock-Johnson Reading Mastery: Word Identification subtest	The Word Identification subtest is a test of decoding skill. It requires the child to read aloud isolated real words that range in frequency and difficulty (as cited in Barker & Torgesen, 1995).

Appendix A3 Summary of findings for the alphabetics domain[1]

Outcome measure	Construct	Study sample	Sample size (students)	Author's findings from the studies: Mean outcome (standard deviation[2]) — *DaisyQuest* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*DaisyQuest* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Level of statistical significance (at α = 0.05)[5]	WWC calculations: Improvement index[6]
Barker and Torgesen, 1995 (randomized controlled trial)									
***DaisyQuest* compared with *Hint and Hunt* software (comparison 1)**									
Undersea Challenge	Phonological awareness	At-risk first graders	49	10.49 (1.10)	9.41 (1.10)	1.08	0.96	Statistically significant	+33
Production Test of Segmenting	Phonological awareness	At-risk first graders	49	7.51 (3.70)	3.27 (2.90)	4.24	1.24	Statistically significant	+39
Phoneme Elision Task	Phonological awareness	At-risk first graders	49	4.38 (3.20)	2.41 (1.70)	1.97	0.74	ns	+27
Sound categorization	Phonological awareness	At-risk first graders	49	9.14 (5.20)	8.43 (4.60)	0.71	0.14	ns	+6
Production Test of Blending	Phonological awareness	At-risk first graders	49	7.30 (4.20)	6.37 (3.50)	0.93	0.23	ns	+9
Phonological awareness average for comparison 1 in Barker & Torgesen, 1995[7]							0.66	ns	+25
Woodcock-Johnson Word Identification subtest	Phonics	At-risk first graders	49	16.35 (9.60)	11.59 (6.40)	4.76	0.57	ns	+22
Analog Reading Task	Phonics	At-risk first graders	49	12.23 (2.40)	12.12 (2.40)	0.11	0.04	ns	+2
Woodcock-Johnson Word Analysis subtest	Phonics	At-risk first graders	49	2.92 (3.30)	1.28 (1.40)	1.64	0.63	ns	+24
Experimental Non-Word Reading	Phonics	At-risk first graders	49	21.84 (9.80)	18.73 (10.60)	3.11	0.30	ns	+12
Phonics average for comparison 1 in Barker & Torgesen, 1995[7]							0.39	ns	+15
***DaisyQuest* compared with math software (comparison 2)**									
Undersea Challenge	Phonological awareness	At-risk first graders	49	10.49 (1.10)	9.40 (0.76)	1.09	1.12	Statistically significant	+37

(continued)

Appendix A3 Summary of findings for the alphabetics domain[1] *(continued)*

Outcome measure	Construct	Study sample	Sample size (students)	Author's findings from the studies: Mean outcome (standard deviation[2]) — *DaisyQuest* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*DaisyQuest* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Level of statistical significance (at α = 0.05)[5]	WWC calculations: Improvement index[6]
Phonics average for comparison 1 in Barker & Torgesen, 1995[7] (continued)									
Production Test of Segmenting	Phonological awareness	At-risk first graders	49	7.51 (3.70)	3.50 (3.90)	4.01	1.03	Statistically significant	+35
Phoneme Elision Task	Phonological awareness	At-risk first graders	49	4.38 (3.20)	2.43 (3.00)	1.95	0.61	ns	+23
Sound categorization	Phonological awareness	At-risk first graders	49	9.14 (5.20)	6.10 (4.60)	3.04	0.60	ns	+23
Production Test of Blending	Phonological awareness	At-risk first graders	49	7.30 (4.20)	5.94 (4.50)	1.36	0.31	ns	+12
Phonological awareness average for comparison 2 in Barker & Torgesen, 1995[7]							0.73	Statistically significant	+27
Woodcock-Johnson Word Identification subtest	Phonics	At-risk first graders	49	16.35 (9.60)	12.39 (8.40)	3.96	0.43	ns	+17
Analog Reading Task	Phonics	At-risk first graders	49	12.23 (2.40)	12.56 (2.30)	–0.33	–0.14	ns	–5
Woodcock-Johnson Word Analysis subtest	Phonics	At-risk first graders	49	2.92 (3.30)	2.03 (2.90)	0.89	0.28	ns	+11
Experimental Non-Word Reading	Phonics	At-risk first graders	49	21.84 (9.80)	19.38 (9.80)	2.46	0.25	ns	+10
Phonics average for comparison 2 in Barker & Torgesen, 1995[7]							0.21	ns	+8
Foster et al., 1994, Experiment 1: Child-care Facility (randomized controlled trial)									
Phonological Awareness Test (PAT) (b)	Phonological awareness	5 year olds	27	22.40 (3.10)	19.20 (3.50)	3.20	0.93	Statistically significant	+32
Screening Test of Phonological Awareness—Experimental Version (STOPA-E)	Phonological awareness	5 year olds	27	18.50 (7.20)	12.40 (6.50)	6.10	0.87	Statistically significant	+31

(continued)

Appendix A3 Summary of findings for the alphabetics domain[1] *(continued)*

Outcome measure	Construct	Study sample	Sample size (students)	Author's findings from the studies: Mean outcome (standard deviation[2]) — *DaisyQuest* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*DaisyQuest* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Level of statistical significance (at α = 0.05)[5]	WWC calculations: Improvement index[6]
Phonological awareness average for Foster et al., 1994, Experiment 1: Child-care Facility[7]							0.90	Statistically significant	+32
Foster et al., 1994, Experiment 2: Kindergarten Classrooms (randomized controlled trial)									
Undersea Challenge	Phonological awareness	Second-semester kindergartners	69	53.30 (9.00)	46.20 (7.40)	7.10	0.85	Statistically significant	+30
Screening Test of Phonological Awareness (STOPA)	Phonological awareness	Second-semester kindergartners	69	26.20 (4.50)	25.30 (7.40)	0.90	0.14	ns	+6
Production Test of Segmenting	Phonological awareness	Second-semester kindergartners	69	11.80 (2.60)	6.00 (4.00)	5.80	1.69	Statistically significant	+45
Production Test of Blending	Phonological awareness	Second-semester kindergartners	69	13.40 (1.90)	10.80 (3.80)	2.60	0.85	Statistically significant	+30
Phonological awareness average for Foster et al., 1994, Experiment 2: Kindergarten Classrooms[7]							0.89	Statistically significant	+31
Mitchell & Fox, 2001 (randomized controlled trial)									
***DaisyQuest* compared with teacher-delivered phonological awareness instruction (comparison 1)**									
Phonological Awareness Test (PAT) (a)—total	Phonological awareness	Kindergartners and first graders	69	73.20 (10.31)	78.30 (11.52)	–5.10	–0.46	ns	–18
***DaisyQuest* vs. other software programs group (comparison 2)**									
Phonological Awareness Test (PAT) (a)—total	Phonological awareness	Kindergartners and first graders	69	73.20 (10.31)	61.60 (16.03)	11.60	0.85	Statistically significant	+30
Domain averages for alphabetics									
All studies							0.62	na	+23
Individual studies									
Barker & Torgesen, 1995[8]							0.52	ns	+20
Comparison 1 in Barker & Torgesen, 1995[8]							0.54	ns	+21
Comparison 2 in Barker & Torgesen, 1995[8]							0.50	ns	+19

(continued)

Appendix A3 Summary of findings for the alphabetics domain[1] *(continued)*

Outcome measure	Construct	Study sample	Sample size (students)	Author's findings from the studies: Mean outcome (standard deviation[2]) — *DaisyQuest* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*DaisyQuest* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Level of statistical significance (at α = 0.05)[5]	WWC calculations: Improvement index[6]
Foster et al., 1994, Experiment 1: Child-care Facility							0.90	Statistically significant	+32
Foster et al., 1994, Experiment 2: Kindergarten Classrooms[9]							0.89	Statistically significant	+31
Mitchell & Fox, 2001[9]							0.20	ns	+8
Comparison 1 in Mitchell & Fox, 2001[10]							–0.46	ns	–18
Comparison 2 in Mitchell & Fox, 2001[10]							0.85	Statistically significant	+30

na = not applicable

ns = not statistically significant

1. This appendix reports overall findings considered for the effectiveness rating and the improvement index. Subtest and subgroup findings from the same studies are not included in these ratings, but are reported in Appendices A4.1 and A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was calculated by the WWC and, where necessary, corrects for clustering within classrooms or schools, for multiple outcomes within one domain, and for multiple comparison groups. For an explanation, see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of *DaisyQuest,* corrections for multiple outcomes and for multiple comparison groups were needed.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. The WWC-computed average effect sizes are simple averages rounded to two decimal places.
8. Values are based on multiple outcomes from two constructs within the alphabetics domain. Furthermore, the overall average for the study is an average across the two comparison groups.
9. Values are based on multiple outcomes from one construct within the alphabetics domain.
10. Values are based on one outcome from one construct within the alphabetics domain. Furthermore, the overall average for the study is an average across the two comparison groups.

Appendix A4.1 Summary of subtest findings for the alphabetics domain[1]

Outcome measure	Construct	Study sample	Sample size (students)	Author's findings from the studies: Mean outcome (standard deviation[2]) — *DaisyQuest* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*DaisyQuest* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Level of statistical significance (at α = 0.05)[5]	WWC calculations: Improvement index[6]
Mitchell & Fox, 2001 (randomized controlled trial)									
***DaisyQuest* compared with teacher-delivered phonological awareness instruction (comparison 1)**									
Phonological Awareness Test (PAT) (a): Rhyming subtest	Phonological awareness	Kindergarteners and first graders	69	17.30 (2.85)	18.20 (3.24)	–0.90	–0.29	ns	–11
Phonological Awareness Test (PAT) (a): Isolation subtest	Phonological awareness	Kindergarteners and first graders	69	22.20 (5.27)	24.70 (3.28)	–2.50	–0.56	ns	–21
Phonological Awareness Test (PAT) (a): Segmentation subtest	Phonological awareness	Kindergarteners and first graders	69	18.70 (3.21)	20.00 (3.32)	–1.30	–0.39	ns	–15
Phonological Awareness Test (PAT) (a): Blending subtest	Phonological awareness	Kindergarteners and first graders	69	15.00 (3.60)	15.40 (4.37)	–0.40	–0.10	ns	–4
***DaisyQuest* compared with other software programs group (comparison 2)**									
Phonological Awareness Test (PAT) (a): Rhyming subtest	Phonological awareness	Kindergarteners and first graders	69	17.30 (2.85)	16.70 (4.29)	0.60	0.16	ns	+6
Phonological Awareness Test (PAT) (a): Isolation subtest	Phonological awareness	Kindergarteners and first graders	69	22.20 (5.27)	15.40 (7.81)	6.80	1.00	Statistically significant	+34
Phonological Awareness Test (PAT) (a): Segmentation subtest	Phonological awareness	Kindergarteners and first graders	69	18.70 (3.21)	16.60 (4.18)	2.10	0.55	ns	+21
Phonological Awareness Test (PAT) (a): Blending subtest	Phonological awareness	Kindergarteners and first graders	69	15.00 (3.60)	12.90 (4.18)	2.10	0.53	ns	+20

ns = not statistically significant

1. This appendix presents subscale findings for measures that fall in the alphabetics domain. Total scores were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was calculated by the WWC and, where necessary, corrects for clustering within classrooms or schools, for multiple outcomes within one domain, and for multiple comparison groups. For an explanation see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of *DaisyQuest,* corrections for multiple outcomes and for multiple comparison groups were needed.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.

Appendix A4.2 Summary of subgroup findings for the alphabetics domain[1]

Outcome measure	Construct	Study sample	Sample size (students)	Author's findings from the studies: Mean outcome (standard deviation[2]) — *DaisyQuest* group	Author's findings from the studies: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*DaisyQuest* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Level of statistical significance (at α = 0.05)[5]	WWC calculations: Improvement index[6]
Foster et al., 1994, Experiment 2: Kindergarten Classrooms (randomized controlled trial)									
Screening Test of Phonological Awareness (STOPA)	Phonological awareness	Second-semester kindergartners	14 lower achieving students (as defined by pretest STOPA)	22.90 (7.30)	17.30 (7.90)	5.60	0.69	ns	+25

ns = not statistically significant

1. This appendix presents subgroup findings for measures that fall in the alphabetics domain. Total group scores were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was calculated by the WWC and, where necessary, corrects for clustering within classrooms or schools, for multiple outcomes within one domain, and for multiple comparison groups. For an explanation, see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of *DaisyQuest,* corrections for multiple outcomes and for multiple comparison groups were needed.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.

Appendix A5 *DaisyQuest* rating for the alphabetics domain

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of alphabetics, the WWC rated *DaisyQuest* as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *DaisyQuest* was assigned the highest applicable rating.

Rating received
Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.
• Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.
Met. *DaisyQuest* had three studies showing statistically significant positive effects, and all of these met WWC evidence standards for a strong design.
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. The WWC analysis found no statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive effects. See the WWC Intervention Rating Scheme for a complete description.