Title
Improving writing skills through technology-based instruction: A meta-analysis

Journal
Review of Education, 6(2)

ISSN
0098-5597

Authors
Little, Callie W
Clark, Jacourie C
Tani, Novell E
et al.

Publication Date
2018-06-01

DOI
10.1002/rev3.3114

Peer reviewed

Review of Education
An International Journal of Major Studies in Education
BERA
BRITISH EDUCATIONAL RESEARCH ASSOCIATION

Review of Education
Vol. 6, No. 2, June 2018, pp. 183–201

DOI: 10.1002/rev3.3114

Improving writing skills through technology-based instruction: A meta-analysis

Callie W. Little[1,*], Jacourie C. Clark[2], Novell E. Tani[3] and Carol McDonald Connor[4]

[1]*University of New England, Armidale, NSW, Australia,* [2]*Florida State University, Tallahassee, Florida, USA,* [3]*Florida Agricultural and Mechanical University, Tallahassee, Florida, USA and* [4]*University of California, Irvine, California, USA*

The present study examined the effect of technology-based writing instruction on writing outcomes using meta-analytic methods. Additionally, this study investigated whether characteristics of study, sample, and outcome moderated the effect of technology-based writing instruction. Six studies were coded resulting in 11 extracted effect sizes. Results revealed that the weighted average effect size for technology-based writing instruction was 0.28, suggesting an educationally relevant and impactful effect of education technology on writing outcomes. Several moderators were included in this meta-analysis, but did not significantly influence effect sizes. One exception was learning disability (LD) status; however, these results should be interpreted with caution as only one study included an LD sample. Overall, these results support previous research and provide knowledge of the populations that are potentially impacted by technology-based writing instruction. Previous literature suggests technology-based writing instruction may supplement teachers' efforts to deliver instruction and provide practice time to students, affording students extra opportunities to engage with writing both in and out of the classroom; however, more research is required to determine the exact mechanisms through which technology may impact writing skills. Recommendations for reporting techniques and directions for future research in development and implementation of technology-based writing instruction are discussed.

Introduction

Students unable to achieve sufficient writing practices are at a disadvantage in comparison with their higher performing peers (Greenwald et al., 1999). Better writing skills enable successful communication of ideas in academic settings as well as in normal day-to-day operations such as composing a social letter to a friend or a professional email to a supervisor or colleague (e.g. Graham et al., 2013). Bearing in mind the importance of writing skills, results of the 2011 National Assessment of Educational Progress (NAEP) revealed that the average writing scores for students in 8th and 12th grades fell below the proficiency level. Only 27% of 8th and 12th graders scored at or above proficiency, with 80% of 8th graders and 79% of 12th graders scoring at or above basic writing levels (NAEP, 2011). Eighth grade students performing at the basic skill level should be able to produce texts that are coherent and effectively

*Corresponding author. School of Psychology and Behavioural Science, S06 Library Road, University of New England, Armidale, NSW 2351, Australia. Email: calliew01@gmail.com.

structured, use supporting details and examples relevant to the main topic, align voice with the topic, purpose and audience, use relevant words and phrases and demonstrate knowledge of spelling, grammar and punctuation with some errors that may impede the meaning of the text. Twelfth grade students performing at the basic level are expected to meet the same requirements as 8th grade students with some errors that do not impede the meaning of the text and the inclusion of appropriately varied simple, compound, and complex sentence types. Students operating at proficient levels in 8th grade should be able to include appropriate connections and transitions, include a variety of simple, complex and compound sentences, demonstrate a solid knowledge of spelling, grammar and punctuation with some errors that do not impede meaning in addition to the basic-level skills. Twelfth grade writers at the proficient level should include all of the basic skills with the addition of more purposefully chosen words, phrases and examples and clearly stated ideas and supporting elements (http://nces.ed.gov/nationsreportcard/writing/achieve.asp).

When an individual fails to develop proficiency in writing, his or her ability to demonstrate knowledge, positions, and philosophies in and outside of school settings is restricted, leading to reduced educational attainment, employment status and quality of life (Graham & Harris, 2005; Graham & Perrin, 2007). In 2002, the last time 4th grade writing was assessed by NAEP, only 28% scored at or above proficient levels, indicating similar trends in performance in elementary, middle and high school students over time (Persky et al., 2003). Technology-based instruction offers the promise of improving these statistics as schools are increasingly implementing technology in the form of online instruction, and computer multimedia to deliver instruction through enriched learning environments (Sarkar, 2012). However, results have been equivocal, with some technology-based interventions reporting large, positive effects (Englert et al., 2007), and others reporting small or negative effects (e.g. Rowley & Meyer, 2003; Goldenberg et al., 2011).

In surveying the literature on technology-based instruction, the range of effects across published studies may be influenced by sample, study and outcome-level moderators. Previous meta-analytic reviews of educational technology have explored moderation of effect size by sample characteristics such as ability level, grade level and socioeconomic status (SES; Cheung & Slavin, 2012) with differential influence of technology-based instruction on students of different ability levels and different grade levels. Additionally, Cheung and Slavin (2012) found significant moderation at the study-level by intervention type and type of research design. The increase in technology-based writing instruction along with the range of effects reported from scientific evaluations of these interventions has led to a growing need to systematically evaluate the efficacy of such programs. The purpose of this meta-analysis is to review research on technology that supports students' writing skills and to examine the efficacy of this technology along with potential moderators of its effect on writing performance.

Background

Writing instruction has been based primarily in one of two main theories of writing development: the 'simple view of writing' and the 'not-so-simple view of writing'.

The simple view of writing separates writing into two basic factors: spelling and ideation. Skilled spelling requires accurately linking orthographic (written) representations of words with their phonological (sound) counterparts and ideation is the ability to generate and organise ideas (Hanna et al., 1966; Juel, 1988). In 2006, Berninger and colleagues updated the simple view of writing by proposing a 'not-so-simple' view of writing that is process-oriented, supported by executive functions and occurs in the working memory system (Berninger & Winn, 2006). Typical writing instruction includes elements that support both the simple and not-so-simple theories of writing development (Cutler & Graham, 2008; Graham & Sandmel, 2011), and a systematic review established evidence across 115 studies that theory-based instructional strategies such as transcription skills (i.e. spelling, typing), strategy instruction, creativity and ideas instruction, self-regulation (i.e. planning, organising), text structure, peer feedback and scaffolding, were effective techniques for improving students' writing skills (Graham et al., 2012). Importantly, Graham and colleagues (2012) also found evidence that the inclusion of a word-processing component to writing instruction improved learning and performance above and beyond traditional writing instruction (Graham et al., 2012).

Despite advances in understanding the writing process and how students learn to write, the NAEP results suggest a need for improved writing instruction in K-12 classrooms. Limitations on how much instructional time teachers can provide to all of the students in their classes may be one barrier between understanding effective writing instruction and successful delivery (Cutler & Graham, 2008; Kellogg & Whiteford, 2009). Cutler and Graham (2008) reported 80% of teachers delivered instruction on basic skills, and process writing strategies at least once a week; however, repetition of previously taught skills was done less frequently, potentially impeding long-term mastery of these skills (Kellogg, 2008; Grabe & Kaplan, 2014). Furthermore, nearly half of teachers reported assigning homework on writing skills or practice between 'never' and 'several times per year' on an 8-point scale of 'never' to 'several times a day' suggesting the majority of children's writing practice was limited to time in class. A 2003 report from the National Commission on Writing recommended doubling the instruction time for writing, providing more out-of-classroom writing time and increasing the use of technology in writing instruction (National Commission, 2003).

Additionally, in 2010, the USA developed the Common Core Standards to provide writing benchmarks for students that are designed to bring students to a proficient level of writing ability at each grade level (National Governors Association, 2010) with emphasis on students' use of technology to produce writing in classroom settings (Graham et al., 2013). Integrating technology into classroom instruction has several hypothesised advantages. For example, a teacher managing a full classroom may be delayed in responding to any one individual, resulting in less overall instructional time and feedback, whereas technology-based programs of writing instruction have the ability to offer feedback and instruction at a student-centred pace (NETP, US Department of Education, 2010). Furthermore, adding technology-based instructional programs into curricula may supplement teachers' instruction by reiterating previously taught skills and providing additional practice opportunities for children in and out-of-classroom settings (Connor et al., 2014).

Integrating technology into instruction has become increasingly easier to implement in the last decade, because of increases in availability of internet services and other resources such as computer memory expansion (Wong & Salahuddin, 2015) and increased processing speed (Khatter & Aggarwal, 2014), and partially because of the concurrent standards and recommendations from the federal government (National Commission, 2003; National Governors Association, 2010). These conditions have provided improved opportunity for development and testing of technology-based writing instruction (Lenhart et al., 2001). In tandem with the growing number of writing technologies, research into their effectiveness has also increased (Rowley & Meyer, 2003; Englert et al., 2007). A 2003 synthesis of technology-based writing has shown a low to moderate, but significantly positive effect on students' writing ability (ES = 0.41, no *p* level reported; Goldberg et al., 2003); however, the majority of studies included in this synthesis implemented word-processing technology that was non-interactive. Additional technological advancements and platforms developed to be utilised in writing instruction have been created since the review conducted by Goldberg and colleagues (2003); newer platforms include advanced features such as multiple skills training modules, interactive tutors and real-time summarisation algorithms (e.g. Rowley & Meyer, 2003; Franzke et al., 2005; Warren et al., 2008). Furthermore, additional research has been conducted on these technology-based interventions using several designs and methods.

The purpose of the current meta-analysis was to review and evaluate these advances in technology-based writing instruction to further inform researchers, educators and practitioners on the efficacy of technology-based writing instruction since 2002. This review did not focus on studies that examined assistive technology, such as word processing or auto-summarisation, in order to examine more directly the link between interactive technology-based writing instruction and writing ability. The following questions were addressed: (1) Does technology-based writing instruction have a significant effect on writing ability among K-12 students? (2) Is the effectiveness of technology-based writing instruction influenced by sources of heterogeneity at levels of sample, study and outcome (e.g. grade level, learning disability status, SES, study strength, skills included in the intervention and type of assessment)?

Method

Search procedure and coding scheme

Figure 1 represents the method of including studies for review. Separate searches were conducted to examine the published literature using PsychInfo, ProQuest, EBSCO and ERIC, and the unpublished literature using OpenGrey, the National Education Policy Center and National Center for Education Statistics databases, and Google Scholar to locate articles published since 2002 that focused on writing instruction provided through a technology-based medium. Studies before 2002 were examined in a previous review (Goldberg et al., 2003) and were not included within the present meta-analysis. Searches were performed using the keywords: 'writing technology instruction', 'computer-adapted writing technology', 'writing technology in education', 'writing technology in classrooms' and 'effectiveness of writing

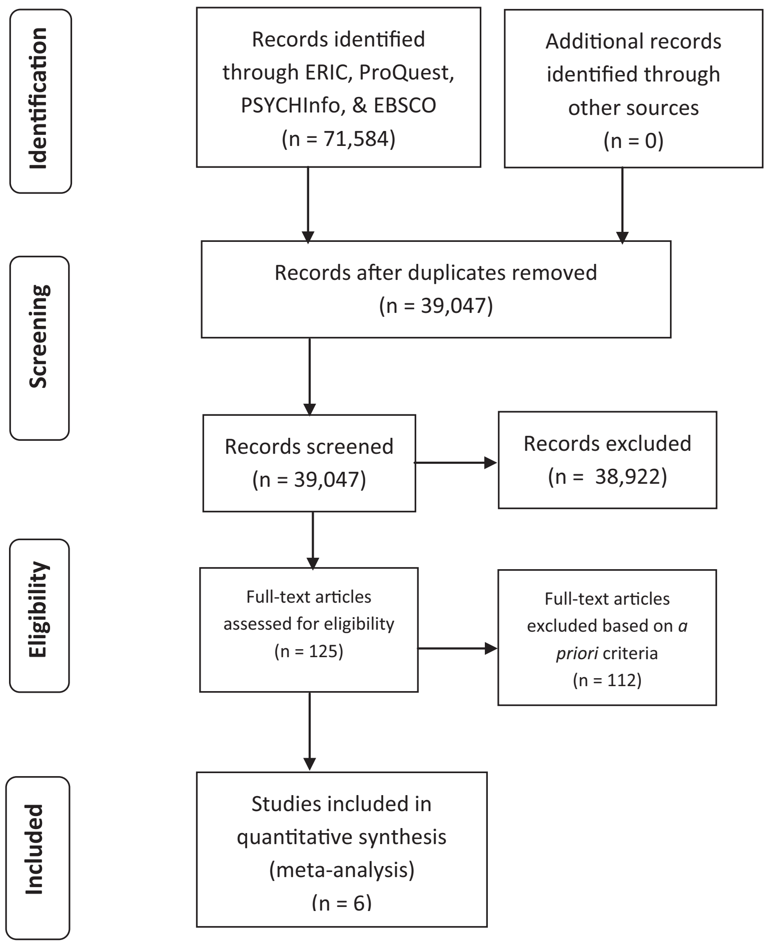

Figure 1. Diagram of study identification, selection, and evaluation for inclusion

technology'. Of the original 71,584 results returned, a review of title, abstract and year of publication led to the exclusion of 71,459. Of the remaining 125 a further 112 were excluded upon review of the entire manuscript. From the initial database results only articles in peer-reviewed, published journals and dissertations were included. The remaining criteria for retaining articles were: (1) the examination of an intervention, (2) a primary analysis of the effectiveness of an intervention, (3) use of a comparison group, (4) the use of a research design that utilised a randomised control design (RCT), quasi-experimental design (QED—must include an equivalent pre-test measure), or regression discontinuity design (RDD); (5) study conducted in English, (6) participants in grades K-12, and (7) study reviewed the outcomes of technology-based interventions on writing skills. In an effort to reduce potential publication bias the authors of the retained studies were emailed requesting any additional data that were either in the manuscript preparation phase or unpublished owing to negative or null effects; however, of the two responses received, no unpublished data were identified which met with the criteria for inclusion. Additionally, references from several technical reports (e.g. *Informing Writing*, *Writing to Read*) were reviewed for studies that met criteria; however, none were located. The final number of articles to code after considering all exclusionary and inclusionary criteria was six. Of the six retained studies, multiple group comparisons within two investigations allowed for a total of eleven separate effect sizes to be extracted.

Studies included were coded at three levels: study, sample and outcome. Study characteristics coded were publication year, experimental design, type of assignment to condition, study strength, sample size of treatment and control as well as overall sample size, type of experimental program or programs (intervention type), the type or types of writing skills included in the intervention and effect size. The assignment to condition variable coded whether participants were assigned through quasi-experimental design, or RCT by group or participant level. The experimental design variable provided additional information on the design by coding whether or not an equivalent pre-test was used between treatment and control groups within the studies. Additionally, sample size was measured at the level of intervention type because two studies included multiple levels of intervention. Study strength was measured using the What Works Clearinghouse (WWC) standards (US Department of Education, 2013): meets WWC evidence standards, meets WWC evidence standards with reservations, or does not meet WWC evidence standards. Intervention type was coded for the type of technology used within the intervention, such as computer-assisted instruction, computer-assisted tutoring or classroom-level technology. The intervention skills variable coded for particular types of writing instruction such as scaffolding, planning, organising text or a combination of these skills. Study-level moderator variables such as experimental design were coded to not only investigate which characteristics of the studies may influence the effect size variability, but also to investigate which levels of these variables are potentially more or less influential. Sample-level variables coded were population (typically performing students or learning disabled). Socioeconomic status (SES; high, medium or low), and participant grade level. The outcome characteristics coded were type of dependent measure (researcher created or standardised) and genre of writing assessment (i.e. narrative or expository). For nearly all studies, writing performance was measured with a holistic score based on a pre-defined rubric. One exception was Rowley and Meyer (2003), which used a teacher-rating scale of how many sessions students needed to obtain writing mastery. Details of the studies and outcome measures are proved in the next section. Table 1 includes all studies with coded moderators (categorical and continuous) and the legend presents the identified categories for each categorical moderator. Additionally, Table 1 presents baseline equivalence, average post-test scores and standard deviations where reported. All variables were initially coded by the first author, and subsequently coded by a trained laboratory assistant to assess inter-coder reliability. The reliability estimates ranged from 0.97 to 0.99 for all coded variables. Table 2 provides a legend for categories appearing in Table1.

Studies coded

Rowley and Meyer *(2003)*. This study evaluated the effectiveness of a Computer Tutor for Writers (CTW); a computer software that assists with the cognitive processes of writing through procedural facilitation. The CTW design focused on the

Table 1. Coded studies

Study	Baseline equivalence	Average post-test score treatment	Average post-test SD treatment	Average post-test score control	Average post-test SD control	Hedges' g	Experi-mental design	Assign-ment	Study strength	Control *n*	Treatment *n*	Technology type	Skills	Assessment type	Assess-ment genre	Pop-ulation	SES	Grade
Rowley & Meyer (2003), G1 vs C	0.09	1.98	0.94	2.25	1.02	−0.18	2	0	1	174	99	1	.	0	3	0	.	8.5
Rowley & Meyer (2003), G2 vs C	0.01	2.36	0.89	2.25	1.02	0.09	2	0	1	174	163	1	.	0	3	0	.	8.5
Rowley & Meyer (2003), G3 vs C	0.55	2.31	1.00	2.25	1.02	0.60	2	0	1	174	36	1	.	0	3	0	.	8.5
Franzke et al. (2005)	0.29	3.32	1.06	2.95	1.30	0.28	1	1	1	59	52	3	4	0	2	0	1	8
Englert et al. (2007)	0.31	3.30	.	2.86	.	1.32	2	0	1	15	20	2	6^1	0	3	2	.	.
Warren et al. (2008)	.	.	.	.	.	0.54	2	0	1	22	22	4	6^2	1	3	.	0	4
Coffman (2011), HTH vs LTH	.	4.45	.	4.23	.	0.40	1	0	2	47	136	5	.	1	1	0	0	11
Coffman (2011), HTH vs LTS	.	4.45	.	4.14	.	0.35	1	0	2	277	136	5	.	1	1	0	0	11
Coffman (2011), LTH vs HTS	.	4.23	.	4.33	.	0.18	1	0	2	47	107	5	.	1	1	0	0	11
Coffman (2011), HTS vs LTS	.	4.33	.	4.14	.	0.34	1	0	2	277	107	5	.	1	1	0	0	11
Goldenberg et al. (2011)	0.21	3.37	0.82	3.49	0.81	0.08	2	0	0	106	193	3	6^3	1	2	0	.	6

Note: Baseline equivalence calculated as the pre-test difference between groups divided by the pooled standard deviation or from *t*-tests. Warren et al. (2008) and Goldenberg et al. (2011) effect sizes calculated from reported *F* statistic. G1 = 2–6 hours of Computer Tutor for Writers, G2 = 6–11 hours of Computer Tutor for Writers, G3 = 11+ hours of Computer Tutor for Writers, C = no training, HTH = high technology use—honors, LTH = low technology use—honors, HTS = high technology use—standard, LTS = low technology use—standard. 6^1 = combination of skills 1, 2, and 3. 6^2 = combination of skills 1 and 5. 6^3 = combination of skills 2 and 3. "." indicates no information was reported in the study. See Table 2 for additional explanation of categories.

Table 2. Legend for Table 1

Category	Value description
Experimental design	1 = Treatment and control, no pretest
	2 = Treatment and control, with pretest
Assignment	0 = Quasi-experimental design
	1 = Randomised control trial by subject
Study strength	0 = Meets WWC evidence standards
	1 = Meets WWC evidence standards with reservations
	2 = Does not meet WWC evidence standards
Technology type	1 = Computer Tutor for Writers (CTW)
	2 = Computer assisted scaffolding
	3 = Computer assisted instruction of summarisation
	4 = Multi-user virtual environment
	5 = Overall classroom technology
Skills	1 = Scaffolding
	2 = Concept mapping
	3 = Organisation
	4 = Summarisation
	5 = Text clues
	6 = Combination of skills
Assessment type	0 = Standardised
	1 = Non-standardised
Assessment genre	1 = Persuasive
	2 = Expository
	3 = Combination of genres
Population	0 = Typical performers
	2 = Learning disabled
SES	0 = Low/economically disadvantaged (as reported in studies)
	1 = Middle class (as reported in study)

following components: '1) a central decision-making and management module called the cognitive tutoring engine; 2) a student record-keeping system used to adapt instruction to the needs of the individual student; 3) teaching resources organized according to an expert writing model; 4) a student interface; and 5) a teacher module'. Participants included 54 classes of 8th and 9th grade English students (n = 471 students) in a quasi-experimental, contrasted groups research design. There were three different treatment groups that each received different levels of the treatment. Group one completed at least two classroom writing sessions [instructed by the teacher] and 2–6 hours of CTW instruction, group two did at least four class sessions and 6–11 hours of CTW and group three did at least six class sessions and at least 11 hours of CTW. Writing performance was measured by the teachers' assessments of how long and how many sessions it took students to demonstrate mastery of the skills taught by the CTW.

Franzke, Kintsch, Caccamise, Johnson and Dooley *(2005)*. Franzke and colleagues investigated a computer tutor, Summary Street®, which provided evaluative feedback on students' written summaries via the Latent Semantic Analysis (LSA) statistical algorithm (Landauer & Dumais, 1997; Landauer, 2002). The Summary Street®

software provided iterative graphical feedback on how closely the student's text covered the main ideas and highlighted areas that needed more adequate coverage. The cycles of feedback and summary submission were repeated until the student's text met the preset requirements for topic summation and length. Participants were randomised into 4 weeks of either treatment (n = 52) or control (n = 59) conditions. The control condition wrote summaries using a word-processing system, but did not receive the graphical feedback through Summary Street® software. Pre-test and post-test writing skills were measured by the standardised writing portion of the Colorado Student Assessment Program (CSAP; Colorado Department of Education, 2000).

Englert, Zhao, Dunsmore, Collings and Wolbers *(2007)*. This study utilised a quasi-experimental, pre-test–post-test design. The participants included 35 elementary-age students with disabilities: 20 in the experimental condition and 15 in the control condition. Participants in both conditions completed a writing sample at the beginning of the study and at the completion of the intervention. The experimental condition used TELE-Web, an online software program that provided several levels of scaffolded assistance in writing such as mapping tools and reminders to include relevant components (i.e. supporting details, topic sentences). Previous research indicated that TELE-Web improved writing performance by providing 'anchors' reminding students about features such as text structure or sentence flow (Englert et al., 2004; Englert et al., 2005). Participants in this condition composed writing samples on a computer and the control group used a similar writing strategy, but in a traditional pencil and paper format. For both groups, the scaffolding techniques used by the teachers were identical, the only difference being that students in the TELE-Web group had access to the mapping and writing tools offered by the software. Both groups spent about 30 minutes doing each daily activity for 3–4 days. The scoring rubric used was developed by Englert (2003) and used the following organisational criteria: (1) introduction to the paper's topic, (2) introduction to the paper's subtopics and categories, (3) adequate depth of subtopical coverage through the inclusion of relevant details, (4) breadth of content coverage through the inclusion of several subtopics that were fairly well developed, (5) conclusion, and (6) overall organisation (introduction, details, and conclusion parts).

Warren, Dondlinger and Barab *(2008)*. The Warren et al. study used a quasi-experimental, pre-test–post-test comparison design to evaluate the efficacy of a three-dimensional learning environment on writing achievement. Participants consisted of 44 students in two fourth grade classrooms. Students were randomly assigned to either the treatment (n = 22) or control group (n = 22) classroom. Students in the treatment condition were taught in a technology-supported learning environment called *Anytown*, a virtual environment that teaches writing skills by scaffolding through character dialogue, feedback from the digital system, and visual and textual clues to facilitate learning activities. The control condition utilised more traditional, face to face instruction. Outcomes were measured with standardised tests by the California Achievement Program and the New Jersey Assessment of Skills and Knowledge.

Coffman *(2011)*. This dissertation study examined technology-enhanced writing instruction in a sample of 11th grade students (n = 567) from a public school district in Southwestern Tennessee. The study based its evaluations on comparing the Tennessee Comprehensive Assessment Program (TCAP; Tennessee Department of Education, 2010) writing scores of students within the study who were taught in high technology use classrooms with students who were taught low technology use classrooms. Comparisons were conducted between level of technology use (low or high) and level of course (honours or standard) with assignment based on teacher reports of the level of technology use in their classrooms and level of course provided by the district manual. Technology use was measured on a 5-point Likert scale with a mean score of 2.5 or greater designated as high technology use and a mean score less than 2.5 designated as low technology use.

Goldenberg, Meade, Cooperman and Midouhas *(2011)*. Goldenberg and colleagues utilised a quasi-experimental, pre-test–post-test design. Participants included 371 students from 17 classes in two different middle schools (115 students came from the comparison school, and 256 from the experimental school). The study evaluated the Writing Matters program that instructs teachers on how to implement digital support for writing process education. Its curriculum includes 'a road-map of lessons, assessment resources to be used by teachers, grade-appropriate writing examples, and a tool for publishing student work once completed'. Teachers at the experimental school administered six Writing Matters units. Writing improvement was measured by two timed writing prompts administered to both the treatment and comparison schools. Each essay was scored using NWP's Analytic Writing Continuum that looks at six characteristics of writing on a scale of one to six. The writing characteristics coded were: ideas/content, structure, stance, sentence fluency, diction and conventions.

Statistical procedure

Effect sizes were calculated as Cohen's d using mean gain scores from pre- to post-test when means and standard deviations were present. When means and standard deviations were not present a related F or t statistic was used (Thalheimer & Cook, 2002). The majority of the studies had small sample sizes; therefore, the effect sizes were corrected for bias using Hedges' formula (Hedges, 1981).

With studies that contribute multiple effect sizes, sample dependency among effect sizes may bias results by decreasing sample variance and increasing the probability of Type I error, or detecting a statistically significant effect when none is present (Borenstein et al., 2009). However, aggregating effect sizes within studies reduces the total number of effect sizes in the model thereby reducing the power to detect significant overall effects. In order to determine whether aggregating the effect sizes would reduce the ability to detect a significant average effect of writing technology, a power analysis was conducted according to established guidelines (Valentine et al., 2010) for power of 0.80. Further, the What Works Clearinghouse has established a minimal effect size of 0.25 to be substantively important in education (US Department of Education, 2013); thus, the power analyses were performed for a minimal detectable

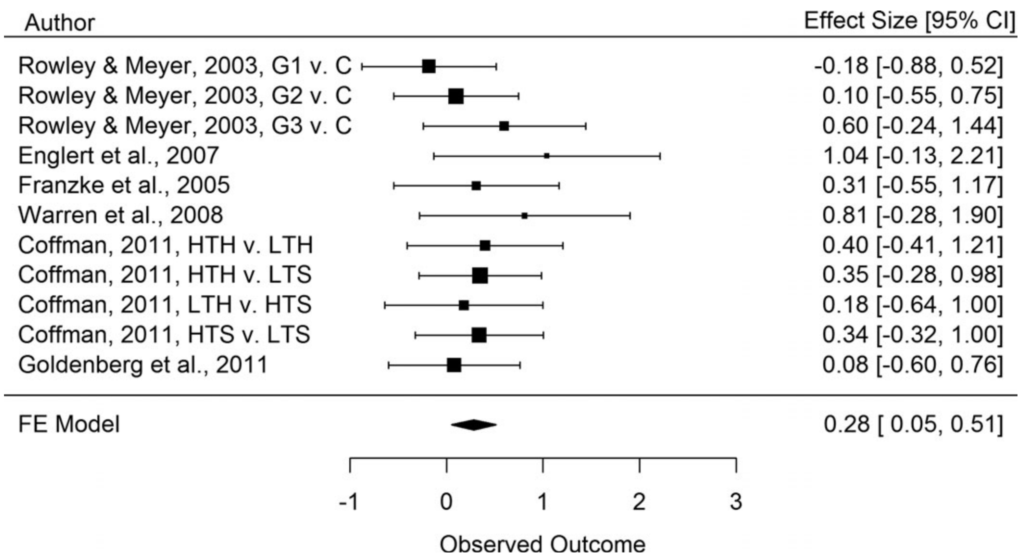

Figure 2. Forest plot of the effect sizes of technology-based writing instruction
Note: G1 = 2–6 hours of Computer Tutor for Writers, G2 = 6–11 hours of Computer Tutor for Writers, G3 = 11+ hours of Computer Tutor for Writers, C = no training, HTH = high technology use—honors, LTH = low technology use—honors, HTS = high technology use—standard, LTS = low technology use—standard

effect size of 0.25. For a fixed effect model, 10 effect sizes were necessary; therefore, we elected to use all available effect sizes within the analyses. All extracted (11) effect sizes and their respective sample sizes were included in fixed effect and random effects models in order to derive the weighted average effect size for technology-based writing intervention (Hedges & Vevea, 1998). After conducting the primary analyses with all 11 effect sizes, follow-up analyses to account for data dependence were conducted using robust variance estimation (RVE; Hedges et al., 2010), which handles dependent effect sizes by clustering the effect sizes by study and weighting them based on correlated effects. RVE provides unbiased estimates of the standard errors absent information on the covariance structure of the effect sizes; therefore, allowing for the inclusion of all effect sizes and eliminating the need for averaging effects.

Moderators were assessed with a two-level, mixed-effects model to predict population-based effect sizes from between study variance (Borenstein et al., 2009). Moderators were examined using Q, I^2, and T^2 statistics. The Q statistic is indicative of either the presence or absence of significant heterogeneity among effect sizes, whereas the I^2 statistic indicates the proportion of variance that is due to heterogeneity versus chance and ranges from 0 to 100% (Q_M; Borenstein et al., 2009). Values of I^2 closer to zero represent variance more likely owing to random error and values closer to 100 represent variance that is more likely owing to true heterogeneity (Higgins et al., 2003). The T^2 values indicate the level of true variance from the observed studies (Borenstein et al., 2009). Finally, potential publication bias was explored using a funnel plot and a Rosenthal fail-safe N test (Cooper et al., 2009). Analyses were conducting utilising the 'metafor' and 'robumeta' packages in R statistical software (Viechtbauer, 2010; R Development Core Team, 2011; Fisher & Tipton, 2015).

Table 3. Moderators of effect sizes for technology-based writing interventions

	n	Q_M (df)	p	T^2	I^2 (%)	R^2 (%)
Continuous predictors						
Year	10	0.29 (1)	0.590	0.051	69.1	0
Grade	10	0.30 (1)	0.582	0.030	59.2	0
Categorical predictors						
Experimental design	11	0.14 (1)	0.706	0.050	68.8	0
Assignment	11	0.00 (1)	0.996	0.056	72.6	0
Study strength	11	0.62 (2)	0.735	0.067	73.1	0
Population	**10**	**6.88 (1)**	**0.009**	**0.029**	**61.7**	**37.4**
Publication type	11	0.14 (1)	0.711	0.51	69.2	0
Assessment type	11	0.01 (1)	0.912	0.056	71.0	0
Assessment genre	11	0.63 (2)	0.729	0.071	74.5	0
Skills	4	0.186 (1)	0.667	0.307	82.1	0
Technology type	11	9.33 (4)	0.053	0.027	58.2	41.2

Note: Significant ($p < 0.05$) moderators are indicated in bold. R^2 indicates the total amount of heterogeneity accounted for by the specified moderator. Not enough studies reported SES for moderator analyses to be conducted.

Results

The result of a fixed-effect test of homogeneity was significant, Q_M (10) = 30.27, $p < 0.01$, indicating heterogeneity between studies. To follow up on the presence of heterogeneity, a random-effects model was conducted, which indicated, in general, technology-based writing instruction has a small, positive and significant effect on writing ability in K-12 classroom environments 0.28 [0.12–0.44], SE = 0.08. Effect sizes ranged from –0.18 to 1.32. Of the 11 comparisons included, 45.5% (5) yielded significantly positive effect sizes, 45.5% (5) yielded non-significant, positive effect sizes, and 9% (1) yielded a non-significant, negative effect size. Figure 2 displays a forest plot of effect sizes and confidence intervals from all studies. Results of the RVE analyses, which accounted for multiple effect sizes from within the same study, indicated a slightly larger effect size of 0.35 [0.02–0.67], SE = 0.12. Sensitivity analyses were conducted across varying values of ρ (0.0, 0.2, 0.4, 0.6 and 0.8) and results indicated the effect sizes (0.34–0.35), standard errors (0.12) and τ^2 values (0.07) were robust to ρ value fluctuations.

In order to determine which features of study, sample and outcome may contribute to heterogeneity, studies were analysed at the moderator level using a mixed model approach. Results of these moderator analyses are presented in Table 3. Publication year and grade were analysed as continuously distributed moderators and the results revealed no significant contribution to heterogeneity between studies from either continuous moderator. The remaining moderators were entered as categorical variables, and results indicated a significant influence on between-studies variability from sample population, alone. No other categorical moderators contributed to heterogeneity between studies. Table 4 indicates the average weighted effect size of technology-based writing instruction for typical performers and for learning disabled populations. Of the studies that reported population characteristics, four studies (nine effect sizes)

Table 4. Effect sizes of technology-based writing instruction by moderator

Population	ES	SE	p	Lower CI	Upper CI
Typical performers	**0.22**	**0.074**	**0.003**	**0.08**	**0.37**
Learning disabled	**1.10**	**0.420**	**0.009**	**0.28**	**1.93**

Note: Significant ($p < 0.05$) effect sizes are indicated in bold.

reported the sample consisted of typically performing students, and one study (one effect size) reported the sample consisted of learning disabled students. Effect sizes were small and significant in studies using a sample of typically performing students; however, the intervention that targeted a learning disabled population produced a large and significant effect size, suggesting the potential for technology-based writing instruction to have a larger impact for those with learning disabilities. For SES, few studies reported levels and of those that did report SES, the majority were homogenous, severely limiting the moderator analysis. Moderation analyses conducted using RVE revealed the same pattern of results, suggesting these results were robust to dependence of effect sizes in two of the studies included in the analyses.

To assess the potential for publication bias, a Rosenthal fail-safe *N* test was conducted. This analysis calculates the number of studies with non-significant results that would be needed to reduce the overall significance of the current results to a non-significant level (Rosenthal, 1979). The results of the fail-safe *N* test indicated 134 studies with null results would need to be added to the meta-analysis to bring the average effect size to a non-significant level. Additionally, a funnel plot was created to further investigate the presence of any potential publication bias. The funnel plot, represented in Figure 3 indicates an asymmetrical distribution of effect sizes, with a gap or absence of effect sizes on the lower left side of the funnel. Effect sizes that are null or negative should fill in the lower left portion of a funnel plot and the absence of points in this area generally suggests some publication bias may be present. However, the fail-safe *N* test suggests that a large number of null or negative studies may be required to nullify the positive significance of technology-based writing instruction found. Lastly, we conducted a trim and fill analysis to estimate the number of 'missing' studies. Results of the trim and fill analysis indicated that three studies were missing from the left side of the funnel plot (see Figure 4). The estimated effect of

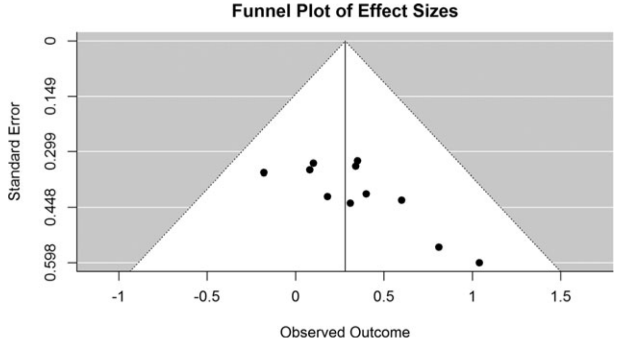

Figure 3. Funnel plot of effect sizes

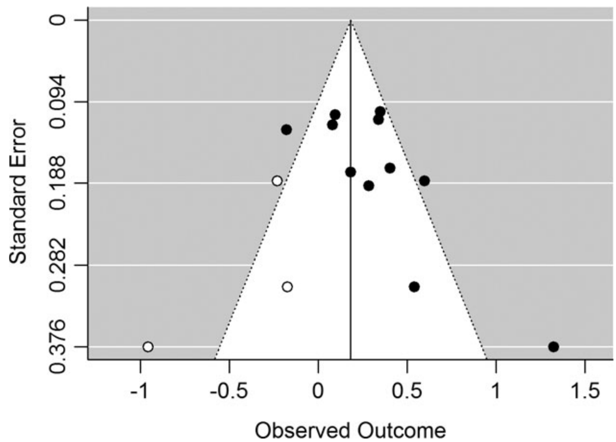

Figure 4. Funnel plot of trim and fill analysis

technology-based writing instruction with the missing studies filled in was smaller in magnitude (ES = 0.19; 95% CI: 0.001–0.376). Although the estimated effect size remained statistically significant, the lower bound confidence interval was barely above zero (0.001) suggesting that publication bias may significantly alter the results of the meta-analysis.

Discussion

The goal of the present meta-analysis was to examine the average effectiveness of technology-based writing instruction on writing outcomes and to investigate whether these effects were moderated by characteristics of study, sample and outcome. Improved understanding of the effectiveness of technology-based writing instruction on writing ability can inform best practices for development and implementation in K-12 educational settings. Results revealed that the weighted average effect size for technology-based writing instruction was 0.28, which, based on the benchmark provided by the WWC and empirically established standards (Hill et al., 2008), suggested an educationally important effect of technology-based writing instruction. Moreover, the current results indicated that learning disability status was a significant moderator of effect size such that the effect size for children with disabilities was greater than for typically developing children (see Table 4); however, other moderators included in this meta-analysis did not significantly influence effect sizes. These findings supported previous studies reporting a positive effect of technology-based writing instruction on writing performance (e.g. Goldberg et al., 2003), extended knowledge of the populations impacted by technology-based writing instruction and provided several recommendations for reporting techniques and future investigations.

The average weighted effect size (0.28) derived from the current set of studies was smaller in magnitude to that of the previous meta-analysis on technology-based writing instruction (0.41; Goldberg et al., 2003), though both indicated a positive and significant effect. One potential explanation for the difference in magnitude may be the focus Goldberg and colleagues directed towards word-processing technology over 'heavily multimedia-enhanced' technology (Goldberg et al., 2003). The present meta-analysis included studies with larger variation in skill instruction and interactivity, with perhaps some types of skill instruction or interactivity having greater influence on writing performance than others. To account for this, the current meta-analysis tested for heterogeneity using several study-level moderators such as type of skill instruction, study strength, study design, and type of technology used; however, results indicated no significant moderation was present. Another potential explanation for the difference in effect size magnitude is that traditional writing instruction has improved and become more standardised in the years since the previous meta-analysis was conducted (National Commission, 2003; National Governors Association, 2010), leading to a smaller overall effect of technology-based writing instruction (Graham et al., 2015) over and above 'business as usual' classroom instruction. Nevertheless, more research is needed to determine how writing performance has changed since the 2011 NAEP assessment. Fortunately, according to the NAEP schedule of assessments, writing will once again be measured in 2017 (https://nces.ed.gov/nationsreportcard/about/assessmentsched.aspx) allowing for an updated assessment of progress. In addition to study-level moderators, sample and outcome-level moderators were tested; but, as previously stated, sample population was the only significant moderator detected. Effect sizes were larger for students with learning disabilities than for children who were reported as typically developing. It is important to interpret this result with caution; however, because only one study within this meta-analysis included an LD population (Englert et al., 2007). Furthermore, Englert and colleagues utilised a quasi-experimental design, meaning the observed effect may have been related to teacher and student characteristics (i.e. years of teaching or student reading level) or factors other than the intervention condition. Finally, although sample population was found to be a significant moderator, only one of the available studies in this meta-analysis examined an LD population. A larger sample of studies with varying conditions or teacher and student characteristics may not replicate the current effect for LD students. More research is needed to determine if the influence of technology-based writing instruction on LD populations is robust, and, if so, which specific elements of instruction steer this influence.

While interesting, the results of the current meta-analysis should be considered with some limitations in mind. First, only a small sample of high-quality manuscripts on technology-based writing instruction were available to include in this review. The authors conducted an extensive and thorough search of scientific databases, the grey literature databases and reached out to researchers for unpublished data; however, the total number of studies to be included was limited by the amount of research that has been produced in this area. Additionally, it is noted that the search did not include cognate terms such as 'text composition' or 'authoring', which may have widened the resulting set of studies to be coded, but were excluded from the search to avoid studies focusing solely on word-processing technology without an interactive component. Given the average weighted effect size obtained from the current results, more investigation of technology-based writing instruction is recommended. A large number of studies failed to provide information on writing skills taught as a part of the intervention and information on sample SES, resulting in low power to detect moderation. With more available information it may have been possible to detect relations between variance in SES levels or skills taught and effect sizes. Lastly, few studies included information on the length or fidelity of intervention implementation although these aspects have been hypothesised to influence the effectiveness of educational interventions (O'Donnell, 2008). Fidelity reports if available, could have broadened the scope of moderator analyses, informed best practices in implementing technology-based writing instruction and formed a basis for improving future efficacy and scale-up studies.

Despite its limitations, the present meta-analysis provides vital contributions to the fields of writing research, educational technology, and special education. This investigation provided an updated systematic review of technology-based writing instruction and revealed that, while still in early phases of practical use, this type of instruction continues to show positive influence on students' writing outcomes across multiple settings, conditions and samples. Technology may supplement teachers' efforts to deliver instruction and practice time to students, affording students extra opportunities to engage with writing both in and out of the classroom (Connor et al., 2014). In fact, several of the included intervention studies described scaffolded and individualised feedback for students, suggesting technology as an effective means of providing student-centred and personalised instruction (e.g. Rowley & Meyer, 2003, Franzke et al., 2005, Warren et al., 2008). Based on these results, future innovators and developers of technology-based writing instruction may wish to further examine the role of scaffolding and other types of student-centred instruction. Moreover, the larger influence of technology-based writing instruction indicated for LD students has implications for which students, educators and administrators may consider prioritising technology-based writing resources towards. If LD students truly benefit from extra time, scaffolding and one-one-one instruction provided by technology, then special education professionals may see greater gains in writing outcomes when technology-based writing instruction is included as part of the standard curricula. While more research is needed to determine the exact mechanisms through which technology supports the acquisition of writing skills, these results provided informative and insightful directions for future research into development and application of technology in writing instruction. To improve the quality of investigation, future studies of technology-based writing instruction should strive to collect and report all available sample characteristics along with detailed information about skill instruction, dosage and fidelity. Through such consideration, research may improve knowledge of within which contexts technology impacts writing outcomes, how these impacts occur and for which populations.

Acknowledgements

The first author was supported by Predoctoral Interdisciplinary Fellowship (funded by the Institute of Education Sciences, US Department of Education, Grant No. 186000-520-025833 and and a National Institute of Health Grant No. P50 HD052120. Views expressed herein are those of the authors and have neither been reviewed nor approved by the granting agencies.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

Berninger, V. W. & Winn, W. (2006) Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution, in: C. A. MacArthur, S. Graham & J. Fitzgerald (Eds) *Handbook of writing research* (New York, Guilford Press), 96–114.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009) *Introduction to meta-analysis* (Chichester, UK, Wiley).

Cheung, A. C. & Slavin, R. E. (2012) How features of educational technology applications affect student reading outcomes: A meta-analysis, *Educational Research Review*, 7(3), 198–215.

*Coffman, M. L. (2011) *The impacts of web-enhanced versus traditional writing instruction*, 1–136.

Colorado Department of Education (2000) *Colorado student assessment program: technical report 2000* (Denver, CO, Author).

Connor, C. M., Goldman, S. R. & Fishman, B. (2014) Technologies that support students' literacy development, in: *Handbook of research on educational communications and technology* (New York, Springer), 591–604.

Cooper, H., Hedges, L. V. & Valentine, J. C. (Eds) (2009) *The handbook of research synthesis and meta-analysis* (New York, Russell Sage Foundation).

Cutler, L. & Graham, S. (2008) Primary grade writing instruction: A national survey, *Journal of Educational Psychology*, 100(4), 907–919.

Englert, C. S. (2003) *Scoring rubrics for evaluating elementary students' written performance*. Unpublished paper, Michigan State University, East Lansing, MI.

Englert, C. S., Manalo, M. & Zhao, Y. (2004) 'I can do it better on the computer.' The effects of technology-enabled scaffolding on emergent writers' personal narrative composition, *Journal of Special Education Technology*, 19(1), 1–15.

Englert, C. S., Wu, X. & Zhao, Y. (2005) Cognitive tools for writing: Scaffolding the performance of students through technology, *Learning Disabilities Research & Practice*, 21(3), 184–198.

*Englert, C. S., Zhao, Y., Dunsmore, K., Collings, N. Y. & Wolbers, K. (2007) Scaffolding the writing of students with disabilities through procedural facilitation: Using an Internet-based technology to improve performance, *Learning Disability Quarterly*, 30(1), 9–29.

Fisher, Z. & Tipton, E. (2015). *Robumeta: An R-package for robust variance estimation in meta-analysis*. arXiv preprint arXiv:1503.02220.

*Franzke, M., Kintsch, E., Caccamise, D., Johnson, N. & Dooley, S. (2005) Summary Street®: Computer support for comprehension and writing, *Journal of Educational Computing Research*, 33(1), 53–80.

Goldberg, A., Russell, M. & Cook, A. (2003) The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002, *The Journal of Technology, Learning and Assessment*, 2(1), 1–51.

*Goldenberg, L., Meade, T., Cooperman, N. & Midouhas, E. (2011) Impact of a technology-infused middle school writing program on sixth-grade students' writing ability and engagement, *Middle Grades Research Journal*, 6(2), 75–97.

Grabe, W. & Kaplan, R. B. (2014) *Theory and practice of writing: An applied linguistic perspective* (Abingdon, Oxon, Routledge).

Graham, S., Gillespie, A. & McKeown, D. (2013) Writing: Importance, development, and instruction, *Reading and Writing*, 26(1), 1–15.

Graham, S. & Harris, K. (2005) *Writing better: Effective strategies for teaching students with learning difficulties* (Baltimore, MD, Brookes).

Graham, S., Harris, K. R. & Santangelo, T. (2015) based writing practices and the common core: Meta-analysis and meta-synthesis, *The Elementary School Journal*, 115(4), 498–522.

Graham, S., Kiuhara, S., McKeown, D. & Harris, K. R. (2012) A meta-analysis of writing instruction for students in the elementary grades, *Journal of Educational Psychology*, 104(4), 879–896.

Graham, S. & Perin, D. (2007) What we know, what we still need to know: Teaching adolescents to write, *Scientific Studies of Reading*, 11(4), 313–335.

Graham, S. & Sandmel, K. (2011) The process writing approach: A meta-analysis, *The Journal of Educational Research*, 104(6), 396–407.
Greenwald, E. A., Persky, H. R., Campbell, J. R., Mazzeo, J., White, S., Vanneman, A., . . . Skinner, R. (1999). NAEP 1998 Writing Report Card for the Nation and the States, *Education Statistics Quarterly*, 1(4), 23–28.
Hanna, P. R., Hanna, J. S., Hodges, R. E. & Rudorf, E. H. (1966). *Phoneme–grapheme correspondences as cues to spelling improvement* (Washington, DC, US Government Printing Office).
Hedges, L. V. (1981) Distribution theory for Glass's estimator of effect size and related estimators, *Journal of Educational and Behavioral Statistics*, 6(2), 107–128.
Hedges, L. V. & Vevea, J. L. (1998) Fixed-and random-effects models in meta-analysis, *Psychological Methods*, 3(4), 486–504.
Hedges, L. V., Tipton, E. & Johnson, M. C. (2010) Robust variance estimation in meta-regression with dependent effect size estimates, *Research synthesis methods*, 1(1), 39–65.
Higgins, J., Thompson, S. G., Deeks, J. J. & Altman, D. G. (2003) Measuring inconsistency in meta-analyses, *British Medical Journal*, 327(7414), 557–560.
Hill, C., Bloome, H., Black, A. R. & Lipsey, M. W. (2008) Empirical benchmarks for interpreting effect sizes in research, *Child Development Perspectives*, 2(3), 172–177.
Juel, C. (1988) Learning to read and write: A longitudinal study of 54 children from first through fourth grades, *Journal of Educational Psychology*, 80(4), 437–447.
Kellogg, R. T. (2008) Training writing skills: A cognitive developmental perspective, *Journal of Writing Research*, 1(1), 1–26.
Kellogg, R. T. & Whiteford, A. P. (2009) Training advanced writing skills: The case for deliberate practice, *Educational Psychologist*, 44(4), 250–266.
Khatter, H. & Aggarwal, V. (2014, February) Efficient parallel processing by improved cpu–gpu interaction, in: A. Sharma, A. Ahlawat, A. Pandey, V. Sharma (Eds) *Issues and challenges in intelligent computing techniques* (ICICT 2014) (Piscataway, NJ, Institute of Electrical and Electronics Engineers), 159–161.
Landauer, T. K. (2002) On the computational basis of cognition: Arguments from LSA, in B. H. Ross (Ed.) *The psychology of learning and motivation* (Vol. 41) (New York, Academic Press), 43–84.
Landauer, T. K. & Dumais, S. (1997) A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge, *Psychological Review*, 104(2), 211–240.
Lenhart, A., Simon, M. & Graziano, M. (2001) *The internet and education: Findings of the Pew Internet & American Life Project* (Washington, DC, Pew Internet & American Life Project).
NAEP. (2011) *National Assessment of Educational Progress: The nation's report card*. National Center for Education Statistics, Washington, DC. Available online at: http://nces.ed.gov/nationsreportcard/
National Commission on Writing in America's Schools and Colleges. (2003) *The neglected "R": The need for a writing revolution*. New York: College Board. Available online at: www.writingcommission.org
National Governors Association & Council of Chief School Officers. (2010) *Common core state standards*. National Governors Association, Washington, DC. Available online at: http://www.corestandards.org
New Jersey Department of Education. (2006) *New Jersey assessment of knowledge and skills*. New Jersey Department of Education, Trenton, NJ. Available online at: http://www.state.nj.us/education/assessment.
O'Donnell, C. L. (2008) Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research, *Review of Educational Research*, 78(1), 33–84.
Persky, H. R., Daane, M. C. & Jin, Y. (2003) The Nation's Report Card: Writing, 2002. National Center for Education Statistics, Washington, DC.

R Development Core Team. (2011) *R: A language and environment for statistical computing* [Computer software]. R Foundational for Statistical Computing, Vienna, Austria. Available online at: http://www.R-project.org

Rosenthal, R. (1979) The file drawer problem and tolerance for null results, *Psychological Bulletin*, 86(3), 638–641.

*Rowley, K. & Meyer, N. (2003) The effect of a computer tutor for writers on student writing achievement, *Journal of Educational Computing Research*, 29(2), 169–187.

Sarkar, S. (2012) The role of information and communication technology (ICT) in higher education for the 21st century, *The Science Probe*, 1(1), 30–41.

Tennessee Department of Education. (2010) *Standards, learning expectations, and performance indicators*. Tennessee Department of Education, LaFollette, TN. Available online at: http://www.tn.gov/education/ci/english/index.shtml

Thalheimer, W. & Cook, S. (2002) How to calculate effect sizes from published research: A simplified methodology. *Work-Learning Research*. Available online at: http://www.work-learning.com/

US Department of Education (2010) *National Education Technology Plan: Transforming American education: Learning powered by technology* (Washington, DC, US Department of Education).

US Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, March). *What Works Clearinghouse: Procedures and standards handbook* (Version 3.0). US Department of Education, Institute of Education Sciences, Washington, DC.

Valentine, J. C., Pigott, T. D. & Rothstein, H. R. (2010) How many studies do you need? A primer on statistical power for meta-analysis, *Journal of Educational and Behavioral Statistics*, 35(2), 215–247.

Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor package, *Journal of Statistical Software*, 36(3), 1–48.

*Warren, S. J., Dondlinger, M. J. & Barab, S. A. (2008) A MUVE towards PBL writing: Effects of a digital learning environment designed to improve elementary student writing, *Journal of Research on Technology in Education*, 41(1), 113–140.

Wong, H. S. P. & Salahuddin, S. (2015) Memory leads the way to better computing, *Nature Nanotechnology*, 10(3), 191–194.