Article

"Tutoring Adolescents in Literacy: A meta-analysis"

Seung Won Jun, Gloria Ramirez et Alister Cumming

McGill Journal of Education / Revue des sciences de l'éducation de McGill, vol. 45, n° 2, 2010, p. 219-238.

TUTORING ADOLESCENTS IN LITERACY: A META-ANALYSIS

SEUNG WON JUN *University of Toronto*
GLORIA RAMIREZ *Thompson Rivers University*
ALISTER CUMMING *University of Toronto*

ABSTRACT. What does research reveal about tutoring adolescents in literacy? We conducted a meta-analysis, identifying 152 published studies, of which 12 met rigorous inclusion criteria. We analyzed the 12 studies for the effects of tutoring according to the type, focus, and amount of tutoring; the number, age, and language background of students; and the quality of the research. Despite variability, these studies suggest benefits, notably for cross-age tutoring, reading, and small tutoring programs of lengthy duration.

PRODIGUER DU TUTORAT EN LITTÉRATURE AUX ADOLESCENTS : UNE MÉTA-ANALYSE

RÉSUMÉ. Qu'est-ce que la recherche nous apprend sur le tutorat en littératie auprès des adolescents? Nous avons mené une méta-analyse, relevant 152 études publiées, parmi lesquelles 12 rencontraient des critères rigoureux d'inclusion. Nous avons donc analysé ces 12 études, examinant les effets du tutorat non seulement selon son type, ses objectifs et sa quantité mais également selon le nombre, l'âge et le profil langagier des élèves. La qualité des travaux de recherche a aussi été prise en considération. Ainsi, malgré une certaine variabilité, ses études suggèrent des bénéfices aux initiatives de tutorat, particulièrement le tutorat inter-âge, les programmes de lecture et les programmes de taille modeste, de longue durée.

Among the many forms of mentoring, one-to-one tutoring may be the most longstanding, conventional, and widely practiced supplement to traditional classroom-based education (Fashola, 2001; Shanahan, 1998). Much research has investigated the benefits of tutoring, particularly for reading but also other school subjects, during the initial years of schooling. Numerous reviews and meta-analyses have synthesized the research on tutoring in elementary schools, establishing clearly its effectiveness: most recently, Ritter, Barnett, Denny and Albin (2009) but also D'Agostino and Murphy (2004), Elbaum, Vaughn, Hughes and Moody (2000), Shanahan (1998), Topping and Hill (1995) and Wasik (1998).

As the National Reading Panel (2000) in the U.S. concluded, early intervention is more effective than remediation later in school. So educators have sought ways to identify young students at risk when there is still time to provide them focused, relevant interventions. Tutoring is generally considered among the most powerful forms of intervention, particularly for increasing the reading achievement of students at risk for academic failure (Burns, Senesec, & Symington, 2004; Elbaum et al., 2000; Harmon, Keehn, & Kenney, 2004; Wasik & Slavin, 1993). As Baker, Gerten, and Keating (2000) observed, "even the best instructional environments for first graders in a public school setting, with one expert teacher responsible for teaching 20-30 students, cannot match the educational intensity of a one-to-one interaction (p. 494)."

Not even small group instruction is as effective as one-on-one tutoring. Ehri, Dreyer, Flugman, and Gross (2007), for example, showed that one-on-one tutoring, rather than small group instruction, was more effective for teaching reading to struggling readers because tutoring allowed instruction to be tailored to the individual needs of student readers, engaging them in greater, focused reading practice with feedback than was feasible in small groups. Likewise, Juel (1996, pp. 268-282) described the characteristics of tutoring that provide advantages over classroom-based teaching, particularly for literacy: Tutors can engage learners with texts and learning processes for concentrated, lengthy periods of time; focus the attention of young learners; model and scaffold reading and writing processes; and provide immediate, individualized feedback in context and other personalized activities at key moments and repeatedly as may be needed.

As this account suggests, tutoring is not a uniform process. Rather, tutoring operates under variable conditions that may be more or less optimal for student learning. For instance, Wasik (1998) and Wasik and Slavin (1993) argued that for literacy tutoring to be effective, (a) tutors need to be supervised by a certified reading specialist, (b) tutors need ongoing training and feedback, (c) tutoring sessions need to be intensive, consistent, structured, and regularly administered, (d) tutors need to use high quality materials, (e) the assessment of tutees needs to be ongoing, and (f) tutoring needs to be coordinated with classroom instruction.

Given the extensive inquiry, positive results, and practical knowledge about tutoring literacy for young learners, it seems surprising that only a limited amount of inquiry has systematically evaluated tutoring for adolescents, and only in recent years. Indeed, as numerous books and reviews have observed, it is only in the past decade that a significant body of research has emerged on literacy development and instruction among adolescent students (Franzak, 2006; Hull & Schultz, 2001; Moje & O'Brien, 2001; Rush, Eakle & Berger, 2007; Schultz & Fecho, 2005). Among this inquiry, claims have started to appear about the benefits of tutoring at-risk adolescent students in literacy in respect

to: discovering the joy and purpose of reading (Cohen, 2007), developing a greater sense of awareness as learners (Friedland & Truscott, 2005), overcoming time constraints in school curricula (Gaffney, Methven, & Bagdasarian, 2002), and mobilizing cross-generational or community resources (Allor & McCathren, 2004; Baker, et al., 2000; Rowen & Gosine, 2006). Two handbooks have also appeared with suggestions for pedagogy and program organization, drawing on educators' practical experiences tutoring adolescents: Chandler-Olcott and Hinchman (2005) and Richards and Lassonde (2009).

But we are not aware of any study, prior to the present one, that has systematically synthesized and evaluated the published research on tutoring adolescent students in literacy. For this reason, we planned the present meta-analysis to address the fundamental question: *How effective is tutoring literacy for adolescent students, according to the results of published research?* Recognizing that conditions for tutoring vary, we also sought to assess the effects of relevant moderator variables, particularly the type, focus, and amount of tutoring; the number, age, and language background of students tutored; and the quality of the research reported.

METHOD

We chose meta-analysis to address the research on literacy tutoring for adolescents in order to synthesize and evaluate the claims made by quantitatively-oriented research. We recognized early on, as Franzak (2006) and Harmon et al. (2004) have shown, that most of the published research on adolescent literacy is, in fact, oriented toward descriptive, case study accounts with qualitative forms of analysis, which vary greatly in respect to educational contexts and the operationalization of key concepts such as tutoring, literacy, at risk, or learning. For that reason, we have conducted a separate, companion analysis (Kohls & Wilson, in press), using methods of content analysis to synthesize further and to critique this research from a complementary perspective. By adopting meta-analysis here we have assumed that different contextual conditions and outcome measures can be standardized in a way that they are statistically comparable across studies, specifically through the calculation of effect sizes. We followed Lipsey and Wilson's (2001) methodological recommendations for meta-analysis, considering empirical studies that present quantitative results that are relatively similar in terms of research design and that are conceptually comparable.

Inclusion criteria

We first collaborated to identify features integral to our search and analysis of studies of tutoring adolescents in literacy, determining that we wanted to include only studies that: (a) were experimental or quasi-experimental in design, (b) involved interventions in the form of tutoring; (c) focused on read-

ing, writing, or other literacy-related skills; (d) reported quantitative outcome measures; (e) provided sufficient information to calculate effect sizes; (f) had a minimum of 10 participants; (g) involved students between 12 and 18 years of age in secondary schools, but not in post-secondary programs; (h) were published within the past two decades (i.e., between 1988 and 2008) in a peer-reviewed journal. In addition to these criteria, we decided to exclude studies that involved participants with learning disabilities, because we were interested in pedagogical implications that could be applied to general populations of learners. To ensure the recognized quality of the research, we limited our search to articles appearing in peer-reviewed scholarly journals, and excluded books, dissertations, and articles published in non-peer reviewed journals, recognizing that this criterion may nonetheless produce a "publication bias" in favor of research with distinct, notable results (rather than non-significant results, which may not tend to be published).

Database Search

We searched the Education Resources Information Center (ERIC) and the Scholars Portal Search databases under the subject area "Social Sciences" with all the specific databases ticked, using different combinations of the following keywords and descriptors: "tutor*," "conferenc*," "adolescen*," "secondary school," "middle school," "high school," "students," "literacy," "reading," "writing," "at risk," "high risk," and "computer-mediated communication." We also searched the reference lists of other meta-analysis studies on tutoring.

This initial search produced a bibliography of approximately 1,800 potentially eligible articles. We entered these citations with abstracts into EndNote, a reference manager software program. We read through the titles of these articles and then their abstracts to identify as many studies as possible relevant to our focus. Upon careful reading, most of the 1,800 studies were excluded because the participants were not adolescents, the research methods were inappropriate, or tutoring was not involved. Only 152 articles met our inclusion criteria. We carefully read their abstracts again, then the full articles if we were unable to decide on suitability from the abstracts alone. After extensive discussion, we selected 23 articles.

We read these 23 articles thoroughly and further identified eight studies that did not strictly meet our inclusion criteria: One did not provide any information on how the tutoring was performed, one reported results for both primary and secondary school students combined (in a way that could not be distinguished), one included students with learning disabilities, one focused on online discussion, and four did not meet the age criterion. This left 15 articles, of which only 12 provided enough information to calculate effect sizes. We wrote to the authors of the three studies that had insufficient details, requesting more statistical information on their research, but we did not receive any responses

from them. Therefore, in the meta-analysis we considered only the 12 studies that provided sufficient statistical information.

Coding study characteristics[1]

We read the final set of 12 studies carefully and repeatedly, coding them according to the design of the study, the characteristics of tutoring (e.g., type, focus, and amount of tutoring), the characteristics of participants (e.g., age, language background, and number of participants), the outcome measures used, the treatment and the control conditions, and the quality of the study. Tables 1 and 2 summarize these characteristics for the 12 studies. Aware that quantitative analyses of such a small number of studies provide limited implications, we also prepared Table 2 to describe the treatment and control conditions as well as the pedagogical significance for tutoring implied, aiming to provide descriptive information that numbers alone would not.

Calculation of effect sizes

We used Cohen's *d* effect size measures. Some studies provided means and standard deviations, from which we could directly calculate effect sizes. Other studies employed different kinds of analytic techniques, presenting results in the form of *t*, *F*, *r*, or χ^2 values, from which we calculated approximate effect sizes. We followed Lipsey and Wilson (2001) for these various methods of calculating effect sizes.

Because Cohen's *d* is based on the mean difference of two groups being compared, for experimental and quasi-experimental studies, we calculated effect sizes based on the mean difference between the experimental and control groups after the tutoring treatment. However, there were three studies (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Hough, Peyton, Geier, & Petrie, 2007; McKinstery & Topping, 2003) for which we had to calculate effect sizes based on the mean gain scores of the treatment group from pretest to posttest. Although Lipsey and Wilson (2001) warned against using different types of effect size statistics in the same comparison, we did not want to further reduce the already small number of studies included for the meta-analysis by excluding these studies; we decided that the advantages of including these three conceptually meaningful studies outweighed the risks of excluding them. Furthermore, pretest differences among comparison groups were corrected for three studies (Franzke, Kintsch, Caccamise, Johnson, & Dooley, 2005; Hough et al., 2007; Penney, 2002), whereas other studies did not provide such information.

TABLE 1. *Summary of study characteristics: Tutoring, participants, quality rating, and effect size*

Study		Tutoring			Participants'			Quality of Study	Effect Size
		Type	Focus	Amount	Age	Language	*N*		
Allen & Chavkin (2004)		adult	reading, math	13.25 - 61 hrs	6th-8th G	18% didn't speak English as L1	746	3.5	0.47
Britt & Aglinskas (2002)	Study1	computer-based	sourcing skills	two 40-minute classes	11th G		15	6.5	1.63
	Study2	computer-based	sourcing skills	two 40-minute classes	11th G		29	6.5	1.02
	Study3	computer-based	sourcing skills	one/two 40-minute classes	11th G		23	6.5	1.71
Chi et al. (2001)	Study1	adult	reading	1.5 - 2 hours	8th G		11	4.5	3.40
	Study2	adult	reading	1.5 - 2 hours	8th G		11	4.5	4.87
Fisher (2001)		cross-age	reading	one school year	7th G	from an ethnically diverse city	45	3.5	1.91
Franzke et al. (2005)		computer-based	reading, writing	two 45-minute classes a week for 4 weeks	8th G	24% from minority groups	111	7	0.20
Hough et al. (2007) Webcam Tutoring		computer-based	reading, writing	two 60-minute sessions a week for 9 weeks	12-19 Y		15	4	1.36
In-person Tutoring		adult	reading, writing	two 60-minute sessions a week for 9 weeks	12-19 Y		10	4	1.18
Jacobson et al. (2001)		cross-age	reading	8 months	7th G	none spoke English as L1	*21	3.5	0.80
McKinstery & Topping (2003)		cross-age	reading,	three 30-minute sessions a week for 9 weeks	12th Y		15	2	0.45
Penney (2002)		adult	decoding skills	fifteen to eighteen 56-minute sessions	Average 17.2 Y	all were Caucasian	33	3.5	1.19
Rowley & Meyer (2003)	G1-Control	computer-based	writing	2 - 6 hours	8th-9th G		273	5.5	-0.19
	G2-Control	computer-based	writing	6 - 11 hours	8th-9th G		337	5.5	0.10
	G3-Control	computer-based	writing	11 or more hours	8th-9th G		210	5.5	0.59
Rowley et al. (1998)	Study2	computer-based	writing	14 hours	8th-9th G		2,018	3.5	0.16
	Study4	computer-based	writing	15 hours	9th G		617	3.5	0.28
Vogelvisesche et al. (2006)		cross-age/ peerage	computer skills	2 hours a week for 11 weeks	Average 15.75 Y	41% were immigrants	207	6	0.34

Note: G1, G2, and G3 = Group1, Group2, and Group3. G = Grade. Y = Year. Shaded areas indicate that the information was not provided in the source article.
*There were 21 participants in the experimental group. The number of participants for the control group was not reported.

TABLE 2. *Summary of study characteristics: Treatment and control conditions and pedagogical significance*

Study	Treatment	Control Group	Pedagogical Significance
Allen & Chavkin (2004)	1 hour a week, 13 50 to 60 hours in total. Focus on homework or on practicing skills such as multiplication or reading comprehension.	Received less amount of tutoring.	Tutoring by minimally-trained tutors was possible to help a student move from a failing to a passing grade.
Britt & Aglinskas (2002) Study1 Study2 Study3	2 class-period exposure to the Sourcer's Apprentice tutorial application.	2 periods of regular classroom activities. Studied the same content, organized as a textbook.	Tutoring sourcing skills for writing was effective. Well-structured computer-based tutoring can be effective.
Chi et al. (2001) Study1 Study2	1.5-2 hours of tutoring. Natural tutoring according to the tutor's preference and style. Tutors suppressed taking a leading role and invited dialogues from students.	No control group. Examined pre-post gains.	The tutors normally took a leading role in a natural setting, but interacting with the students was also shown to be effective.
Fisher (2001)	After the teacher modeled the lessons, the tutors practiced and implemented the lessons. They had discussions and journal writing later.	Attended a remedial reading class.	Providing students with an authentic reason for literacy, regular feedback, and modeling and integrating writing into the curriculum were effective.
Franzke et al. (2005)	2 sessions a week for 4 weeks. Wrote summaries and received feedback from the Summary Street computer tutor.	Did not receive feedback on their summaries from Summary Street.	Guided practice and feedback that directs students to attend to relevant content was important in summary writing.
Hough et al. (2007) Webcam Tutoring In-person Tutoring	18 one hour sessions. A session consisted of repeated reading, summaries, word study, guided reading, writing, and reading aloud.	No control group. Examined pre-post gains for the two groups of tutoring.	Well-structured web-cam tutoring can be effective, gives students easier access to tutoring, and lower their anxiety about being tutored.

TABLE 2. *Summary of study characteristics: Treatment and control conditions and pedagogical significance (Continued)*

Study	Treatment	Control Group	Pedagogical Significance
Jacobson et al. (2001)	After the teacher modeled the lessons, the tutees practiced and implemented the lessons. They had discussions and journal writing later.	Students from another middle school who did not have the strategic reading class.	Well-structured tutoring with guidance and opportunity for practice and reflection was effective.
McKinstery & Topping (2003)	3 sessions a week for 4 weeks. Pair reading and Paired Thinking, which embeds the teaching of thinking skills in the transferable skill of reading.	No control group. Examined pre-post gains.	Well-structured tutoring with modeling and scaffolding was effective. Tutoring in thinking skills was an effective way of improving reading comprehension.
Penney (2002)	18 sessions. Students read aloud, then practiced decoding skills for difficult words.	Remained in the classroom.	Tutoring decoding skills to improve word reading was effective.
Rowley & Meyer (2003) G1-Control G2-Control G3-Control	Different amount of tutoring. Students received computer tutoring that guided them through the writing process.	Received regular classroom instruction.	Well-structured computer tutoring, based on a cognitive theory of writing, with guidance and scaffolding throughout the writing process was effective.
Rowley et al. (1998) Study2	55 min each, total of 8 hours. Students received tutoring from a user-adaptive computer application.	Used word processor for an equivalent amount of time.	Well-structured computer tutoring, based on a cognitive theory of writing, with guidance and scaffolding throughout the writing process was effective.
Study4		Attended regular English classes without treatment.	
Vogelwiesche et al. (2006)	2 hours a week for 11 weeks. The students received tutoring on basic computer skills from older adolescents who passed the course or from adults.	Other adolescents who did not participate in the computer course.	With guidance, adolescents can be good tutors. The opportunity of tutoring can increase knowledge and self-esteem.

TABLE 3. *Summary of effect sizes*

Study	Outcome Measures	*n*	ES	*ES*	*ES'*	*SE*	*w*
Allen & Chavkin (2004)	pass/fail rate	246	0.42	0.42	0.42	0.13	60.14
Britt & Aglinskas (2002)							
Study1	sourcing performance	15	1.73	1.73	1.63	0.60	2.81
Study2	sourcing performance	29	1.02	1.02	0.99	0.41	5.90
Study3	sourcing performance	23	1.87	1.71	1.65	0.49	4.14
	essay quality		1.54				
Chi et al. (2001)							
Study1	knowledge piece	11	3.53	3.53	3.40	0.67	2.25
	mental model						
	question answering						
Study2	knowledge piece	11	5.06	5.06	4.87	0.85	1.39
	mental model						
	question answering						
Fisher (2001)	vocabulary	45	2.16	1.91	1.88	0.36	7.81
	reading comprehension		1.81				
	SAT-9		1.77				
Franzke et al. (2005)	overall	111	0.11	0.20	0.20	0.19	27.50
	C summary		0.38				
	S inference		0.09				
	A fact-finding		0.22				
	P vocabulary		0.14				
	other		0.17				
	summary quality		0.31				
Hough et al. (2007)							
Webcam Tutoring	reading	15	1.96	1.40	1.36	0.41	6.09
	vocabulary		0.84				
In-person Tutoring	reading	10	1.24	1.23	1.18	0.48	4.26
	vocabulary		1.22				
Jacobson et al. (2001)	SDRT	[a]21	0.80	0.80	0.78	0.32	9.75
McKinstery & Topping (2003)	reading comprehension	15	0.46	0.46	0.45	0.37	7.32
Penney (2002)	word identification	33	1.38	1.19	1.16	0.39	6.61
	word attack		0.88				
	passage comprehension		1.32				
Rowley & Meyer (2003)	writing quality	273	0.19	0.19	0.19	0.13	62.84
G1-Control							
G2-Control	writing quality	337	0.10	0.10	0.10	0.11	84.06
G3-Control	writing quality	210	0.59	0.59	0.59	0.19	29.11
Rowley et al. (1998)							
Study2	writing quality	2,018	0.16	0.16	0.16	0.05	476.84
Study4	writing quality	617	0.28	0.28	0.28	0.08	149.17
Vogelwiesche et al. (2006)	achievement test	207	0.40	0.34	0.34	0.22	21.02
	self-esteem		0.28				

Note. G1, G2, and G3 = Group1, Group2, and Group3. *n* = Total number of participants, ES = effect size, *ES* = average effect size, *ES'* = adjusted effect size, *SE* = standard error, *w* = weight. SAT= Stanford Achievement Test. CSAP = Colorado Student Assessment Program. SDRT = Stanford Diagnostic Reading Test. Shaded areas indicate that effect sizes were not calculable for these measures. [a]The sample size of the control group is not given; 21 were in the experimental group.

Table 3 summarizes the effect sizes for all studies we analyzed. For each study, we calculated an effect size for each outcome measure in the study. When more than one effect size was calculated from a study, we established the average effect size for the study overall. Although we recognized the differences in some outcome measures within a study, conceptually we treated them as a general construct of literacy (e.g., reading, writing, vocabulary) and used all outcome measures to arrive at one average effect size for each study. Therefore, each study contributed one average effect size (*ES*) to the subsequent analyses with the following exceptions: (1) Britt and Aglinskas (2002) consisted of three independent sub-studies, so we calculated an effect size for each sub-study, resulting in three effect sizes; (2) Chi et al. (2001) included two separate sub-studies, resulting in two effect sizes; (3) Hough et al. (2007) compared in-person and webcam tutoring – two different types of tutoring – so we calculated separate effect sizes for each type, resulting in two effect sizes based on pretest-posttest gain scores; Rowley and Meyer (2003) compared three experimental groups with the control group, each involving different amounts of tutoring, so each of the three experimental-control group comparisons produced an effect size; Rowley, Carlson, and Miller (1998) included four independent sub-studies, only two of which had sufficient information to calculate effect sizes.

For studies with 20 or less participants (i.e., Britt & Aglinskas, 2002; Chi et al., 2001; Hough et al., 2007; McKinstery & Topping, 2003), we used Hedge's (1981) adjusted effect sizes, which correct for an upward bias when the sample size is small.

Moderator variables

To examine the effects of tutoring in relation to different mediating factors, we grouped the studies according to the following moderator variables:[2]

(1) Type of tutoring: cross-age peer tutoring, adult tutoring, computer-based tutoring

(2) Focus of tutoring: reading, writing, other literacy skills

(3) Amount of tutoring: 7 hours or less, 8 to 15 hours, 16 hours or more

(4) Participants' age: 6th to 7th grade, 8th to 9th grade, 10th grade or higher

(5) Number of participants: 20 or less, 21 to 70, 71 or more

(6) Studies including second-language (L2) learners

(7) Quality of study: low, medium, high

To account for the different sample sizes in the studies, we established the weight of each effect size as the inverse of the square of standard error (1/

SE^2) of the effect size. As a result, a larger weight was given to effect sizes calculated from studies with larger populations, thereby contributing more to the mean effect size. Furthermore, we used a fixed effects model to conduct the moderator analyses.

Quality indicators

Assuming that results from studies with stronger methodological design provide more accurate and valid information, we accounted for the quality of each study included in the current meta-analysis by applying Graham and Perin's (2007) nine quality indicators, following their standards (p. 452):

(1) Assignment of participants: Studies were classified as involving random assignment of participants to conditions, matching participants without random assignment, and non-random assignment without matching.

(2) Mortality equivalence: Mortality equivalence was met if most of the students starting the study completed it, producing equivalent mortality across conditions.

(3) No ceiling/floor effects at posttest: Ceiling or floor problems were not evident if the mean of the posttest quality measure for each condition was more than one standard deviation away from the lowest and highest score of the scale.

(4) Pretest equivalence: Pretest equivalence was met if the study provided evidence that the performance of students in each of the conditions was equivalent prior to the start of instruction.

(5) Instructor training: Instructor training occurred if there was a description of how teachers were prepared to administer the experimental treatment.

(6) Type of control condition: Control conditions were classified as an alternative treatment that was clearly described or an unspecified or no-treatment control condition.

(7) Hawthorne effect: Hawthorne effect was not evident if the researcher(s) put into place conditions to control for it.

(8) Treatment fidelity: Treatment fidelity was established if evidence was provided that the experimental treatment condition was administered as intended.

(9) Teacher effects controlled: Teacher effects were controlled if tutors were randomly assigned to conditions or if they taught each condition.

We gave a score of 1 for each quality indicator met in each study we analyzed and a score of 0 for those not met. For the "assignment of participants" we gave a score of 1 if the study employed random assignment, a score of 0.5 if

the study did not employ random assignment but matched participants, or a score of 0 if the study neither employed random assignment nor matched participants. Overall, studies that received a total score of less than four, four to six, and greater than six were categorized as being low, medium, and high quality studies, respectively.

The criteria outlined by Graham and Perin (2007) to assign scores on the quality indicators provided a useful base to help achieve consistency. However, our decisions on what score to assign involved a certain level of subjectivity, and the given criteria were not always directly transferrable to the studies included in the current meta-analysis. To reduce bias and increase reliability, one of the authors conducted the initial coding to assign quality scores on each of the nine indicators. Subsequently, issues in relation to specific indicators were discussed at research meetings with senior researchers and Ph.D. students working together on a project focused on adolescent literacy. After a second round of scoring, the first and second authors met to determine and resolve various decisions for allocating points on ambiguous indicators.

RESULTS

Table 4 summarizes our results. The results should be interpreted with caution because a test of homogeneity prior to the moderator analyses indicated that the distribution of effect sizes was not normal. Table 4 shows both the unweighted and weighted mean effect sizes, though we refer to the weighted mean effect size to interpret the findings. All the effect sizes are statistically significant, all are positive, and the range of their confidence intervals do not contain zero. The overall weighted mean effect size for the 12 studies is 0.26, a relatively low magnitude. That is, collectively, adolescents who received literacy tutoring performed 0.26 standard deviations higher than the norm in each educational context. Chi et al. (2001) included two independent sub-studies, each of which produced extremely high effect sizes (3.40 and 4.87). Although outliers, we retained these results in order to preserve as many studies as possible in the meta-analysis.

Type of tutoring[3]

Cross-age tutoring had the highest mean effect size (1.05), followed by adult tutoring (0.70) and computer-based tutoring (0.19). For cross-age tutoring, Fisher (2001) and Jacobson et al. (2001) studied struggling adolescents as tutors, obtaining a higher effect size than did McKinstery and Topping (2003), which investigated students as tutees. These results suggest benefits for at-risk adolescents tutoring younger students. Similarly, adult tutoring of at-risk adolescents proves beneficial. These findings corroborate impressions previously highlighted in the literature, emphasizing the value of cross-age peer tutoring (e.g., Bernstein, Boquiren, & Cho, 1997). The mean effect size for computer-

based tutoring is low (0.19), despite all of the computer-based tutoring studies showing statistically significant results favoring the computer-based tutoring group over control groups. The large sample populations in these computer-based studies probably contributed to the significance in their results.

TABLE 4. *Summary of results for 12 studies*

Moderators		*k*[a]	Unweighted Mean ES	Weighted Mean ES	95% Confidence Interval	
					Upper	Lower
Overall		19	1.13	0.26	0.19	0.32
Type of Tutoring	cross-age	3	1.05	1.05	0.65	1.44
	adult	5	2.21	0.70	0.48	0.93
	computer-based	10	0.69	0.19	0.13	0.26
Focus of Tutoring	reading	9	1.71	0.92	0.69	1.14
	writing	6	0.21	0.17	0.10	0.24
	other skills	4	1.18	0.73	0.40	1.07
Amount of Tutoring	7 hrs or less	7	1.81	0.24	0.05	0.43
	8 - 15 hrs	5	0.32	0.20	0.13	0.27
	16 hrs or more	7	1.03	0.66	0.48	0.84
Participants' Age	6th - 7th grade	4	0.90	0.60	0.39	0.82
	8th - 9th grade	8	1.18	0.18	0.11	0.25
	10th grade or higher	7	1.20	0.90	0.63	1.18
Number of Tutored Participants	20 or less	8	1.95	1.43	1.09	1.76
	21 - 70	5	0.94	0.66	0.44	0.88
	71 or more	6	0.19	0.17	0.11	0.24
Studies with L2 Learners		5	0.73	0.48	0.31	0.65
Quality of Study	low	6	0.84	0.42	0.29	0.55
	mid	8	1.43	0.22	0.10	0.25
	high	5	0.98	0.50	0.25	0.75

[a] Number of effect sizes included in the analyses.

Focus of tutoring[4]

The mean effect size is highest for tutoring reading (0.92), followed by tutoring other literacy skills (0.73), then by tutoring writing (0.17). The magnitude of the mean effect size for reading is quite large, suggesting consistency in the effectiveness of tutoring for students who are struggling to read. Tutoring also appears to be effective for other literacy skills such as using source references and learning computer skills. The low mean effect size for writing here may have arisen because all of the studies considered for tutoring writing employed computer-based tutoring, which, as observed above, produced low mean effect sizes compared to other types of tutoring. So this result may be restricted to computer-based tutoring of writing (cf. Graham & Perin, 2007). Or it may be that analytic rating scales for assessing writing abilities are not sufficiently fine-grained to reveal students' achievements within brief durations or a single course (Cumming, 2003).

Amount of tutoring

The mean effect sizes for small (7 hours or less) and medium (8 to 15 hours) amounts of tutoring is 0.24 and 0.20, respectively. However, for longer durations of tutoring (16 hours or more), the mean effect size is fairly high (0.66). Evidently, time is an important factor in tutoring adolescents. As with other types of instruction, more instruction appears to result in more learning. Alternatively, there may be a minimum, threshold duration needed for tutoring to be effective with adolescents. Many factors could be involved. For example, Friedland and Truscott's (2005) study of tutoring at-risk adolescents emphasized that it took a long time to build positive relations between tutors and tutees.

Participants' age

The mean effect size is highest for students in the 10th grade or higher (0.90), followed by 6th to 7th grade students (0.60), and then 8th to 9th grade students (0.18). The reasons for this trend are not clear. Most of the students at grades 8 and 9 came from two studies (Rowley, Carlson & Miller, 1998; Rowley & Meyer, 2003), which examined computer-based tutoring and produced small effect sizes with very large weights. Otherwise, tutoring seems to be quite effective for adolescents. Lauer et al.'s (2006) meta-analysis, which included K-12 students, also showed the highest effect size for high school students (grades 9 to 12). A less favored possibility is that 8th and 9th graders are affected by maturational and psycho-emotional issues that prevent them from reaping the benefits of tutoring to the same extent as younger and older adolescents.

Number of participants

The mean effect size is the highest for studies with 20 or fewer students (1.43). Overall, mean effect sizes decrease for studies that include more participants.

The mean effect size for studies with 71 or more students is only 0.17. This result implies that tutoring adolescents may be more effective when administered to small numbers of students. When tutoring takes place on a large scale, and a tutor has to be responsible for many students at one time, many of the integrally attractive characteristics of tutoring, such as individualized and focused instruction, might be lost. Even with a sufficient number of tutors, logistical concerns may appear regarding the management of teenagers and the administration of a large-scale program.

Studies with L2 learners

Most of the studies failed to describe the language and cultural backgrounds of participating students, particularly whether the research included (or excluded) L2 learners as participants. Based on the descriptions of the studies, we could identify with certainty only five studies (Allen & Chavkin, 2004; Fisher, 2001; Franzke et al, 2005; Jacobson et al., 2001; Vogelwiesche, Grob, & Winkler, 2006) that included L2 learners. None actually compared groups of L1 and L2 learners. The mean effect size for these five studies is 0.48, a higher number than the overall mean effect size of 0.26 for all 12 studies together, suggesting that tutoring can be effective for L2 learners.

Quality of study

Did studies of higher quality produce greater effects? High quality studies did yield the largest effect size: The mean effect size was highest for high quality studies (0.50), followed by low quality (0.42) and then medium quality studies (0.22). However, the effect size for low quality studies was also as high. With such a small number of studies to compare, it is not clear whether studies that met rigorous quality standards were able to produce greater effects. Considering that quality is to some extent an elusive dimension and relative to the indicators used to evaluate it, it is possible that the quality indicators we used were not sensitive enough to clearly capture and distinguish quality among the studies.

DISCUSSION

The dozen research studies that we examined were diverse. Each had investigated the effects of tutoring different aspects of literacy with adolescent students, mostly reading and writing, but also using source references, computer skills, and reading-related skills such as decoding and high-level thinking. Moreover, the studies involved different types of tutoring: cross-age tutoring, adult tutoring, and computer-based tutoring. The studies also involved differing durations of tutoring, ranging from less than seven hours to a whole academic year. They involved students at different grades in schools and in numbers ranging from 11 to over 2,000. The quality of the research designs and methods also varied in terms of the nine quality indicators we applied. In sum, these studies were

heterogeneous, as was evident in the result of a test of homogeneity. Obviously, 12 studies are a small sample. For these reasons, the results of our meta-analysis merely point toward tendencies rather than general conclusions.

Nonetheless, the tendency is that quantitatively oriented research suggests that tutoring literacy can be effective for adolescents. Distinct effects appear particularly for cross-age and adult tutoring, for tutoring reading, for tutoring that lasts longer, and for tutoring older adolescents. Tutoring also appears to be most effective in studies with high quality research designs, for small numbers of students, and for programs that include L2 learners. So rather than asking "Is tutoring effective?" these results suggest that educators should ask, "How can tutoring be effective?" and organize tutoring programs that capitalize on these factors for adolescent students.

Complementing these results, the pedagogical significance of tutoring implied in the studies (described in Table 2) offer insights into how tutoring can be effective in various educational contexts. In general, the studies suggest that:

- providing students with scaffolding and feedback in tutoring sessions can be effective (Franzke et al., 2005; Jacobson et al., 2001; McKinstery & Topping 2003);
- focusing on specific subskills of literacy, such as decoding skills (Penney, 2002) and thinking skills (McKinstery & Topping, 2003) for reading and sourcing skills for writing (Britt & Aglinskas, 2002), can be beneficial for developing literacy skills;
- tutoring by minimally-trained tutors, including adolescents, can be promising when guidance is provided (Allen & Chavkin, 2004; Fisher, 2001; Jacobson, et al., 2001; Vogelwiesche et al., 2006);
- computer tutoring, if well-structured and informed by theory, can effectively help students that have limited access to face-to-face tutoring (Britt & Aglinskas, 2002; Franzke et al., 2005; Hough et al., 2007; Rowley & Meyer, 2003; Rowley, Carlson & Miller, 1998);
- providing students with genuine reasons to use literacy and with a sense of responsibility by allowing them to tutor other students can help them actively engage in literacy practices (Fisher, 2001; Jacobson et al., 2001; Vogelwiesche et al., 2006); and
- the role of the tutor in a tutoring session shapes learning outcomes (Chi et al., 2001).

The greatest limitation on our meta-analysis was the lack of high quality empirical studies. Many publications did not provide sufficient information to calculate effect sizes or lacked in other aspects of methodological rigor. This limitation forced us to make compromises, such as using effect sizes calculated

from pretest-posttest gain scores, to retain as many studies as possible for the meta-analysis. More research needs to be done and be carefully designed and documented. For future research, we recommend long-term, longitudinal studies with large sample sizes that examine the complexity, processes, and effectiveness of tutoring. To be effective, tutoring programs need to be extensive, and the parameters of optimal duration need to be explored programmatically, especially in view of issues such as the time and means to develop rapport and continuity among tutors and tutees. In addition to the existing focus on reading, more inquiry should address adolescent students' writing, computer-based programs (as new technologies and pedagogical programs appear), differences between culturally and linguistically diverse or homogeneous student populations, and variations in tutor qualifications such as cross-age and volunteer tutoring.

ACKNOWLEDGMENT

We gratefully acknowledge funding for this research from the Social Sciences and Humanities Research Council of Canada, standard grant 410-2006-2442, as well as useful advice on the present analyses from Esther Geva, Olesya Falenchuk, Robert Kohls, and Jennifer Shade Wilson.

NOTES

1. We did not establish inter-coder reliability when coding the study characteristics, calculating the effect sizes, or scoring the quality indicators. Instead, we worked collaboratively and reached full agreement through discussion. There were many unclear cases that required complex decisions.
2. A test of homogeneity showed that the distribution of the effect sizes across the studies cannot be explained from sampling error alone. Despite this constraint, we decided to conduct the moderator analyses, knowing that the results could not be generalized.
3. We did not include Vogelwiesche, Grob, and Winkler (2006) in this analysis because they compared two types of tutoring – same-age tutoring versus cross-age tutoring – for which we were unable to calculate separate effect sizes. In Table 3, effect sizes for Vogelwiesche, Grob and Winkler (2006) are based on the mean differences between the same-age group, treating it as the experimental group, and the cross-age group considered as the control group.
4. Distinguishing reading from other outcomes was complicated in several studies. Allen and Chavkin (2004) had to be excluded from this analysis because, although it involved tutoring both reading and math, the outcome measures did not differentiate reading from math, so it was impossible to distinguish students' achievements in either type of ability. In contrast, Franzke, Kintsch, Caccamise, Johnson, and Dooley's (2005) study of tutoring reading and writing provided results on separate measures of these abilities, so we could calculate effect sizes for each ability distinctly. We categorized McKinstery and Topping's (2003) study of tutoring thinking skills to help students to comprehend texts as tutoring reading. Penney (2002) studied tutoring of decoding skills and included reading outcome measures, so we also categorized it as tutoring reading. Hough et al. (2007) studied tutoring reading and writing together, but they did not include any outcome measures for writing, so we only considered their measures for tutoring reading here.

REFERENCES

* indicates studies included in the meta-analysis

* Allen, A., & Chavkin, N. F. (2004). New evidence that tutoring with community volunteers can help middle school students improve their academic achievement. *School Community Journal, 14*, 2, 7-18.

Allor, J., & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disability Research & Practice, 19*, 116-129.

Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly, 35*, 494-519.

Bernstein, S., Boquiren, C., & Cho, A. (1997). *Keeping cross-age tutoring alive: Growing and sustaining a school-wide tutoring program*. San Francisco, CA: Partners in School Innovation.

*Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*, 4, 485-522.

Burns, M. K., Senesac, B. V., & Symington, T. (2004). The effectiveness of the HOSTS Program in improving the reading achievement of children at-risk for reading failure. *Reading Research and Instruction, 43*, 87-103.

Chandler-Olcott, K. & Hinchman, K. (2005). *Tutoring adolescent literacy learners: A guide for volunteers*. New York: Guilford Press.

*Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science: A Multidisciplinary Journal, 25*, 471-533.

Cohen, J. (2007). A case study of a high school English-language learner and his reading. *Journal of Adolescent & Adult Literacy, 51*, 164-175.

Cumming, A. (2003). Assessing L2 writing: Alternative constructs and ethical dilemmas? *Assessing Writing, 8*, 73-83.

D'Agostino, J. V., & Murphy, J. A. (2004). A meta-analysis of Reading Recovery in United States schools. *Educational Evaluation and Policy Analysis, 26*, 23-38.

Elbaum, B., Vaughn, S., Hughes, M. T., Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*, 605-619.

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading Rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal, 44*, 414-448.

Fashola, O. (2001). *Building effective afterschool programs*. Thousand Oaks, CA: Corwin.

*Fisher, D. (2001). Cross age tutoring: Alternatives to the reading resource room for struggling adolescent readers. *Journal of Instructional Psychology, 28*(4), 234-240.

Franzak, J. K. (2006). "Zoom": A review of the literature on marginalized adolescent readers, literacy theory, and policy implications. *Review of Educational Research, 76*, 2, 209-248.

*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.

Friedland, E. S., & Truscott, D. M. (2005). Building awareness and commitment of middle school students through literacy tutoring. *Journal of Adolescent & Adult Literacy, 48*, 550-562.

Gaffney, J. S., Methven, J. M., & Bagdasarian, S. (2002). Assisting older students to read expository text in a tutorial setting: A case for a high-impact intervention. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 18*, 119-150.

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

Harmon, J. M., Keehn, S., & Kenney, M. S. (2004). Tutoring struggling adolescent readers: A program investigation. *Reading Research and Instruction, 44*, 46-74.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

*Houge, T. T., Peyton, D., Geier, C., & Petrie, B. (2007). Adolescent literacy tutoring: Face-to-face and via webcam technology. *Reading Psychology, 28*, 3, 283-300.

Hull, G., & Schultz, K. (2001). Literacy and learning out of school: A review of theory and research. *Review of Educational Research, 71*, 4, 575-611.

*Jacobson, J., Thrope, L., Fisher, D., Lapp, D., Frey, N., & Flood, J. (2001). Cross-age tutoring: A literacy improvement approach for struggling adolescent readers. *Journal of Adolescent & Adult Literacy, 44*, 528-536.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31*, 268-289.

Kohls, R., & Wilson, J. S. (in press). A thematic synthesis of research on literacy tutoring. To appear in A. Cumming (Ed.), *Understanding adolescent literacy in a culturally diverse context*. New York: Routledge.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*, 275-313.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*McKinstery, J., & Topping, K. J. (2003). Cross-age peer tutoring of thinking skills in the high school. *Educational Psychology in Practice, 19*, 199-217.

Moje, E., & O'Brien, D. (Eds.). (2001). *Constructions of literacy: Studies of teaching and learning in and out of secondary schools*. Mahwah, NJ: Erlbaum.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Rockville, MD: National Institute of Child Health and Human Development Clearinghouse.

*Penney, C. G. (2002). Teaching decoding skills to poor readers in high school. *Journal of Literacy Research, 34*, 1, 99-118.

Richards, J. C., & Lassonde, C. A. (Eds.). (2009). *Literacy tutoring that works: A look at successful in-school, after-school, and summer programs*. Newark, DE: International Reading Association.

Ritter, G., Barnett, J. Denny, G. & Albin, G. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research, 79,* 3-38.

Rowen, N. & Gosine, K. (2006). Support that matters: A community-based response to the challenge of raising the academic achievement of economically vulnerable youth. In D. Armstrong & B. McMahon (Eds.), *Inclusion in urban educational environments: Addressing issues of diversity, equity, and social justice* (pp. 277-299). Charlotte, NC: Information Age Publishing.

*Rowley, K., Carlson, P., & Miller, T. (1998). A cognitive technology to teach composition skills: Four studies with the R-WISE writing tutor. *Journal of Educational Computing Research, 18*, 259-296.

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

Rush, L., Eakle, J., & Berger, A. (Eds.). (2007). *Secondary school literacy: What research reveals for classroom practice*. Urbana, IL: National Council of Teachers of English.

Schultz, K., & Fecho, B. (2005). Literacies in adolescence: An analysis of policies from the United States and Queensland, Australia. In N. Bascia, A. Cumming, A. Datnow, K. Leithwood & D. Livingstone (Eds.), *International handbook of educational policy, Vol. 2* (pp. 677-694). Dordrecht, Netherlands: Springer.

Shanahan, T. (1998). On the effectiveness and limitations of tutoring in reading. In P. D. Pearson & A. Iran-Nejad (Eds.), *Review of research in education* (pp. 217-234). Washington, DC: American Educational Research Association.

Topping, K., & Hill, S. (1995). University and college students as tutors for schoolchildren: A typology and review of evaluation research. In S. Goodlad (Ed.), *Students as tutors and mentors* (pp. 13-31). London, PA: Kogan Page.

*Vogelwiesche, U., Grob, A., & Winkler, B. (2006). Improving computer skills of socially disadvantaged adolescents: Same-age versus cross-age tutoring. *Learning and Instruction, 16*, 241-255.

Wasik, B. A. (1998). Using volunteers are reading tutors: Guidelines for successful practices. *The Reading Teacher, 51*, 562-570.

Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28*, 178-200.

SEUNG WON JUN is a Ph.D. candidate at the Ontario Institute for Studies in Education, University of Toronto. He is interested in second language writing, reading, and vocabulary, and particularly helping English learners to develop appropriate literacy skills in one-on-one tutoring contexts. He received his M.A. in TESOL from Michigan State University.

GLORIA RAMÍREZ is Assistant Professor in the School of Education, Thompson Rivers University, Canada, where she teaches courses on language and literacy development and effective classroom practices. Her research interests are bilingual and biliteracy development in children and adolescents as well as assessment and literacy teaching for minority groups and disadvantaged populations.

ALISTER CUMMING is Professor and Head of the Modern Language Centre at the Ontario Institute for Studies in Education, University of Toronto. His research and teaching focus on writing in second languages, literacy and assessment in classroom and formal testing contexts, and curriculum evaluation, particularly of programs for English as a second or foreign language.

SEUNG WON JUN est doctorant à l'Institut d'études pédagogiques de l'Ontario de l'Université de Toronto. Il s'intéresse à l'apprentissage de l'écriture, de la lecture et du vocabulaire en langue seconde, accordant un intérêt particulier pour l'aide à apporter aux adolescents dans le développement de compétences littéraires appropriées dans un contexte de tutorat individualisé. Il a obtenu son diplôme de maîtrise en enseignement de l'anglais langue seconde à l'Université de l'État du Michigan.

GLORIA RAMÍREZ est professeur associée à l'École des sciences de l'éducation de l'Université Thompson Rivers au Canada. Elle y prodigue des cours sur le développement du langage et de la littératie ainsi que sur les pratiques pédagogiques éprouvées en classe. Ses domaines de recherche portent sur le développement bilingue et bilittéraire chez les enfants et les adolescents ainsi que sur l'évaluation et l'enseignement de la littérature au sein des groupes minoritaires et des populations défavorisées.

ALISTER CUMMING est professeur et Directeur du Centre d'études des langues vivantes de l'Institut d'études pédagogiques de l'Ontario de l'Université de Toronto. Ses recherches et son enseignement ciblent l'écriture des langues secondes, la littératie et son évaluation en classe et/ou dans des contextes formels ainsi que l'évaluation des programmes, particulièrement ceux portant sur l'enseignement de l'anglais comme langue seconde ou langue étrangère.