International Electronic Journal of Elementary Education
Vol. 2, Issue 1, October, 2009.

Calibration of comprehension and performance in L2 reading

Seda SARAC*
Yildiz Technical University, Istanbul, Turkey

Betul TARHAN
Yildiz Technical University, Istanbul, Turkey

Abstract

Comprehension monitoring is crucial for successful reading. Although the researchers appreciate the importance of comprehension monitoring in L2 reading, there are only a few studies done on the comprehension monitoring ability of L2 readers. The main aim of this study was to investigate the comprehension monitoring abilities of university students while reading expository texts in L2. The results showed that the students' were not able to calibrate their comprehension at above chance level whereas they were able to calibrate their performance. The results were discussed comparative to findings from earlier research in L1 reading.

Keywords: metacomprehension, second language, calibration

Introduction

Metacognition refers to any "any knowledge or cognitive activity that takes as its object, or regulates, any aspect of any cognitive enterprise" (Flavell, 1979). There are two facets of metacognition identified by many researchers namely, metacognitive knowledge and metacognitive skills (Baker & Brown, 1984; Veenman, 2005; Veenman & Elshout, 1995). Metacognitive knowledge is what we know about the operations of our cognition (Flavell, 1979; Pintrich, 2002). This knowledge allows us to contemplate our planning, goal setting, processing of tasks, monitoring of progress, and recognition and

* Corresponding author. E-mail for correspondence: ssarac@yildiz.edu.tr

ISSN:1307-9298

www.iejee.com

repair of problems (Grabe & Stoller, 2002). Metacognitive skills are series of activities that support the individuals in controlling their own learning processes (Schraw, 2002). Metacognitive monitoring, that refers to a person's on-line awareness of his or her own cognitive processes (Dunlosky & Lipko, 2007; Pieschl, 2009) is an important metacognitive skill (Nietfeld, Enders, & Schraw, 2006; Pieschl, 2009). Metacognitive monitoring enables learners to assess their performance and use appropriate fix-up strategies in case of failure (Schraw & Moshman, 1995). Accurate metacognitive monitoring is crucial for successful reading (Cromley, 2005; Dunlosky & Lipko, 2007; Wiley, Griffin, & Thiedei, 2005; Zhao & Linderholm, 2008).

Comprehension Monitoring

Comprehension monitoring, which is a form of metacognitive monitoring (Baker & Brown, 1984), is the readers' awareness of the degree to which they understand what they are reading. If readers believe that they have understood the text well, there is no reason for them to go on processing. However, if the reader believes that he fails to comprehend, this awareness leads to the reprocessing of the text. Therefore, monitoring of comprehension is a prerequisite for the effective implementation of comprehension strategies (Kimmel & MacGinitie, 1984). Inaccuracy in comprehension monitoring judgements leads to uncorrected errors in comprehension. Learners with poor comprehension monitoring will not be able to use their judgements to correct their errors and guide their learning appropriately (Nietfeld, Enders & Schraw, 2006).

The three paradigms, widely used by the researchers studying metacognitive monitoring in reading are error detection (e.g., Kolic-Vehovec & Bajsanski, 2006; Otero, Campanario & Hopkins, 1992; Zabrucky & Moore, 1994), calibration of comprehension and calibration of performance (e.g., Dunlosky & Rawson, 2005; Glenberg & Epstein, 1985; Lin & Zabrucky, 1998; Miesner & Maki, 2007; Lin, Zabrucky & Moore, 2002). In error detection paradigm, subjects read a text containing syntactic, lexical and/or semantic errors, and their comprehension monitoring behaviour is assessed by their ability to detect these inconsistencies (Otero, Campanario, & Hopkins, 1992).

In calibration of comprehension paradigm, students read a text and then are asked to make metacognitive judgements before answering the questions (e.g. Epstein, Glenberg, & Bradley, 1984; Glenberg & Epstein, 1985). Among the metacognitive judgement types that are mostly studied are; task difficulty or ease of learning judgements (EOL), learning and comprehension monitoring or judgements of learning (JOL), and confidence judgements (CJ). In EOL, the calibration of comprehension is the relation between easiness and performance. In JOL, it is the relation between understanding level and performance. In case of CJ, it is the relation between predicted performance and actual performance (Lin, Moore & Zabrucky, 2001). In calibration of performance paradigm, students read a

text, answer the questions related to the text and judge how sure they are about the accuracy of their answers. Calibration of performance is the relation between performance judgements and actual performance (e.g. Nietfeld, Cao, & Osborne, 2005; Zabrucky, Agler, & Moore, 2008). In the present study two of these paradigms, *calibration of comprehension* and *calibration of performance*, were used.

Comprehension Monitoring in L1 Reading

Comprehension monitoring in L1 reading is a widely researched topic. Early studies, using calibration paradigms, showed that readers were not able to judge their comprehension levels accurately. Glenberg and colleagues reported that their readers' comprehension monitoring judgements are far from being accurate. Gamma correlations between confidence judgements and performance scores were never higher than .20 and only differed from zero by chance. Readers often overestimate how much they have comprehended (Glenberg & Epstein, 1985, 1987; Glenberg, Sanocki, Epstein, & Morris, 1987). In a study by Weaver and Bryant (1995), it was reported that under certain conditions Gamma correlation coefficient reached .35 level. Maki (1995) reported that Gamma correlations between reading performance and comprehension judgements of adult readers were never higher than 0.27. In 36 different studies carried out in Dunlosky laboratory, similar results were obtained (Dunlosky & Lipko, 2007). Lin, Moore and Zabrucky (2001) assessed their students' calibration performance using the students' pre-test judgements of understanding, confidence, easiness and interestingness. The researchers reported that the students were able to calibrate their comprehension at above chance level. The mean Gamma correlation coefficients were .15, .14, .14 and .14 for understanding judgements, confidence judgements, easiness judgements and interestingness judgements respectively. Although the students were able to calibrate their comprehension at above chance level, their calibrations were still very low.

With regard to the relation between comprehension monitoring and reading performance, research results suggest a weak link between these variables. According to Cavanaugh and Perlmutter (1982) and Pressley and Schneider (1997) there is no strong empirical evidence linking monitoring accuracy to measures of reading comprehension. In a study by Begg, Martin, and Needham (1992), the relation between accuracy and test performance were investigated. They found that participants that less accurately monitored their learning were more successful than participants that more accurately monitored their learning. Several studies by Dunlosky and his colleagues (Dunlosky & Connor, 1997; Connor, Dunlosky & Hertzog, 1997) had similar results. Their research showed that the groups that differed in performance did not differ in monitoring accuracy. In a study by Lin, Moore and Zabrucky (2001), students' understanding, confidence, easiness, and interestingness judgements did not correlate with their reading performance, indicating that good comprehenders were not necessarily good

monitors, or vice versa. In contrast, Metcalfe (2009) asserted negative correlations between accuracy of metacognitive monitoring judgements and study time allocation found in most studies in their laboratories. According to Metcalfe (2009) accurate metacognitive monitoring judgements are prerequisites of successful learning. Accurate metacognitive monitoring judgements lead to successful learning but only with the provision that the learners/readers are able to use this information to determine their study behaviours i.e. implementing appropriate strategies. In a similar vein, Thiede, Anderson and Therriault, (2003) study indicated that learners with accurate monitoring judgements study more strategically and become more successful.

The Present Study

Although metacognition is a research area deemed important by L2 reading researchers, most studies are limited to pedagogical interventions and strategy use reading (Morrison, 2004). Comprehension monitoring in L2 is rather a "neglected essential", as stated by Casanave (1988), for L2 reading researchers.

According to Casanave (1988), comprehension monitoring in L2 reading is neglected because most L2 reading studies are schema theory-driven. According to the schema theory, people adjust their memories of a culturally unfamiliar story to fit a "schema" that is more consistent with their own culturally familiar knowledge of the typical content and structure of stories. This theoretical view caused L2 reading researchers to deal more with the effects of content and structure of texts on reading comprehension of L2 readers and to neglect what the readers do while trying to comprehend the text in L2.

Block (1992), who underscored the importance of comprehension monitoring in L2 reading, asserted various reasons for this importance. Firstly, L2 readers may be able to reflect on their cognitive processes. This awareness, then, brings about more appropriate judgements when reading in L2 than reading in L1. Secondly, reading in L2 is more difficult than reading in L1 as L2 readers encounter more unfamiliar language and need more awareness of the reading processes in order to use appropriate fix-up strategies when they experience comprehension failure.

Not more than a few studies can be found on the metacognitive monitoring processes of readers while reading texts in L2. It can also be seen that there are only a few studies done on the comprehension monitoring ability of L2 readers and that the researchers have used only the error detection paradigm in those studies.

What has especially compelled the researchers of the present study to further investigate this subject is that little is known about the calibration accuracy in L2 reading, while much more is known about L1 reading. Therefore, the first aim of this study is to examine students' accuracy of calibration of comprehension and calibration of performance in L2 reading.

For the calibration of comprehension, three indices were used; namely, Judgement of Learning (JOL), Ease of Learning (EOL) and pre-test Confidence judgements (PreCJ). For the calibration of performance, students' post-test Confidence judgements (PostCJ) were used. The second aim of the study is to investigate the intercorrelations between different calibration measures. The third aim of the study is to examine the relationship between L2 readers' metacognitive knowledge and their calibrations.

Method

Participants

Participants were 42 undergraduate TEFL (Teaching English as a Foreign Language) students at Yildiz Technical University, Turkey. The mean age of the participants was 20.95 year (*SD* = .31). All the participants spoke Turkish as their first language and English as their second language. Most of the students (*n* = 38) started to learn English at the age of 12. The remaining students' (*n* = 4) starting age was seven. 31 out of 42 students could speak a third language other than Turkish and English. None of the students have the experience of living in a country other than Turkey. Of all the students, only 1 student visited an English-speaking country. All the students volunteered for the study.

Materials

Students' metacognitive knowledge was assessed by The Metacomprehension Scale (Moore, Zabrucky & Commander, 1997). The Metacomprehension Scale (MCS) consisted of 22 statements under seven subscales. Agreement with each statement was indicated on a 5-point scale (1=disagree strongly - 5= agree strongly). The seven subscales were Regulation (methods of resolving comprehension failures), Strategy (techniques to improve reading), Task (knowledge of basic comprehension processes), Capacity (perception of comprehension abilities), Anxiety (stress related to comprehension performance), Locus (control of reading skills) and Achievement (importance of good comprehension skills). Cronbach's alpha reliability coefficient is .77 for this study, indicating a reasonably reliable measure of metacognitive knowledge for this sample.

The reading texts and comprehension questions used for this research were taken from standardized YDS examinations for the study. The YDS (abbr. for *Yabancı Dil Sınavı*) examinations, designed to test some certain aspects of linguistic competence in either of the three languages, English, French, German, are taken by the candidates for the BA programmes offered by Turkish universities in the fields of the literature or the teaching of a language other than Turkish. The examination is designed and administered by the *OSYM*, a testing centre coordinating a number of standardized tests in Turkey.

The texts were chosen from the YDS for two reasons primarily. Among our major concerns was designing a valid and reliable testing instrument to be utilized in the research. YDS exam items are piloted for validity and reliability via statistical tools and can be confidently used for the purpose of the research. Our second concern was providing a testing instrument that is levelled suitably for the target group. A suitable testing instrument must have the linguistic level not any higher than the participants'. The participants of the this study had already taken an English test of a similar difficulty level to be admitted for the university BA program, English Language Teaching, which lead the researchers to the assumption that the YDS item difficulty is suitable for the subjects to perform the required monitoring processes.

Eight single paragraph expository texts were chosen for this study. The longest text was 195 words and the shortest was 115 (M = 170; SD = 26.1). Flesch-Kincaid Reading Ease scores ranged between 33. 4 and 68.2 (M = 52.1; SD = 11). The students' text comprehension performance was assessed by four inference questions for each text.

Three prediction-rating scales for calibration of comprehension (JOL, EOL, and PreCJ) and one postdiction scale for calibration of performance (postCJ) were prepared. The first prediction scale (JOL) referred to how well the subjects think they understand the text and ranged from 1, designating "not at all", to 4, designating "very well". The second prediction scale (EOL) referred to how easy the subjects find the text and ranged from 1, designating "not easy at all", to 4, designating "very easy". The third prediction scale (preCJ) referred to how certain the subjects feel about their answers to the inference questions about the text and ranged from 1, designating "not sure at all", to 4, designating "very sure". The Postdiction scale (post-CJ) referred to how certain they are that they answer all the questions correctly and ranged from 1 designating "not at all sure" to 4 designating "very sure".

Procedure

The students were tested in a group session. The entire session took approximately two hours. Texts were distributed in a booklet form. The order of the texts was the same for all the participants. The students were given three minutes to read each text. After the students read each text, they were asked to complete three sets of Likert-type scales. When the students finished their task with the scales, the marked scales were collected from the students so that they would not make any changes later. This step was followed by the distribution of the comprehensions questions about the texts read to the students. The students were allowed to refer to the texts as they were answering the questions. The entire session took approximately an hour. The second session took place one day after the first session. The students filled out The Metacomprehension Scale. The entire session took approximately half an hour.

Results

Descriptive Statistics

Metacognitive Knowledge. Students' metacognitive knowledge scores, assessed by Metacomprehension Scale, ranged between 54 and 1001. Mean metacognitive knowledge score for 42 students was 74.14 (*SE* = 1.44).

Comprehension Performance. Students' local and global comprehension performance scores were determined. Local comprehension performance scores were based on the total number of questions the students answered correctly out of 4 multiple choice questions for each text. Mean local comprehension scores for 42 students were for 2.98 (*SE* = .16) for text one; 2.60 (*SE* = .17) for text two; 3.38 (*SE* = .12) for text three, 2.36 (*SE* = .16) for text four; 3.07 (*SE* = .10) for text five; 3.17 (*SE* = .14) for text six; 2.55 (*SE* = .18) for text seven; and 3.48 (*SE* = .10) for text eight. Global comprehension performance scores were based on the total number of multiple-choice questions the students answered correctly out of 32 multiple choice questions. Mean global comprehension score for 42 students was 23.60 (*SE* = .65).

Calibration Measures. Students' calibration of comprehension and calibration of performance were determined by correlating students' ratings for each text and their actual performance on that text. To find out the relations between students' ratings and their actual performances, Gamma correlations were used. Gamma (G) is a non-parametric correlation that requires ordinal data and recommended for the data of this type (Nelson, 1984). Gamma is a symmetric measure of association. It ranges from -1 (if higher ratings are always paired with lower performance) and +1 (if higher ratings are always paired with higher performance). Zero correlation indicates that there is no correlation between variables. Students' comprehension performance was based on the total number of multiple-choice questions they answered correctly.

With the purpose of determining the calibration of comprehension, three contingency tables were prepared for each student. The first table contained the student's JOL rating for each text (ranging from 1 to 5) and local performance score for each text (ranging from 0 to 4). The second table contained the student's EOL rating for each text (ranging from 1 to 5) and local performance score for each text (ranging from 0 to 4). The third table contained the student's CJ rating for each text (ranging from 1 to 5) and local performance score for each text (ranging from 0 to 4). Gamma correlations were computed to determine the relations between each scale rating and the local performance scores. The average Gammas for the three judgements were calculated. The mean JOL, EOL and CJ gammas for this study were .140 (*SE* = .092), .046 (*SE* = .086) and .113 (*SE* = .085)

respectively. One-sample t-tests for JOL, EOL and CJ gammas were also computed to investigate whether these Gamma correlations were beyond chance levels, that is, if they were significantly different from zero. The results were ($t(41) = 1.519$, $p > .01$), ($t(41) = .540$, $p > .01$) and ($t(41) = 1.332$, $p > .01$) for JOL, EOL and CJ respectively. None of results was significant, that is, these results indicated that the students were not able to calibrate their comprehension at above chance levels.

With the purpose of determining the calibration of performance, a contingency table was prepared for each student containing post-test CJ rating for each text (ranging from 1 to 5) and local performance score for each text (ranging from 0 to 4). Gamma correlations were computed to determine the relations between the scale ratings and the local performance scores. The average Gamma for post-test CJs was computed. The mean post-test CJ Gamma in this study was .51(SE= .056). One-sample t-test for post-test CJs showed that this Gamma was significantly different from zero ($t(41) = 9.132$, $p > .01$) indicating that the students were able to calibrate their performance at above chance level. These results indicated that students in this study were not able to calibrate their comprehension whereas they were able to calibrate their performance at above chance level.

Correlations

Interrelations among Calibration Measures. Pearson product-moment correlation coefficients were computed to investigate the interrelations among JOLs, EOLs, pre-test CJs, post-test CJs Gamma coefficients. The results are presented in Table 1. All the calibration of comprehension measures (JOL, EOL, and PreCJ) are correlated significantly and positively with each other whereas the calibration of comprehension measure (postCJ) has only significant correlation with EOL judgements and the correlation is negative. These results indicate that the students with poor EOL calibration are good at calibrating their past performance whereas the students with good EOL calibration are poor at calibrating their past performance (See Table 1).

Calibration Measures and Comprehension Performance. Pearson product-moment correlation coefficients were computed to investigate the correlations between each calibration measure and global comprehension performance. Global comprehension performance (GCP) is only correlated significantly with EOLs and the correlation is negative, that is, students with good EOL calibration are poor comprehenders whereas students with poor EOL calibration are good comprehenders. The other calibration measures did not significantly correlate with global performance scores (See Table 1).

Calibration Measures and Metacognitive Knowledge. Pearson product-moment correlation coefficients were computed to investigate the correlation between calibration measures and metacognitive knowledge. The results showed that metacognitive knowledge is only correlated significantly with

postCJs, that is, students with high levels of metacognitive knowledge are better performance calibrators than students with low metacomprehension knowledge. No significant correlations are found between comprehension calibrations and metacognitive knowledge (See Table 1).

Table 1. Correlations among Calibration Measures, General Comprehension Performance and Metacognitive Knowledge

	Calibration of Comprehension			Calibration of Performance		
	JOL/C	EOL/C	PreCJ/C	PostCJs	GCP	MK
JOL/C	/					
EOL/C	.42**	/				
PreCJ/C	.74**	.45 **	/			
PostCJ/C	-.19	-.47**	-.19	/		
GCP	-.13	-.34*	-.21	-.04	/	
MK	-.24	-.26	-.11	.31 *	-.11	/

Note: JOL/C= Calibration based on JOLs, EOL/C= Calibration based on EOLs, PreCJ/C= Calibration based on pre-test CJs, PostCJ/C= Calibration based on post-test CJs.
* $p < .05$; ** $p < .01$

Discussion

The accuracy of metacognitive judgements in L1 reading is a widely researched topic among reading researchers. However, in case of L2 reading, there are few studies conducted to investigate the accuracy of readers' metacognitive judgements.

Major goal of this study is to investigate accuracy of L2 readers' metacognitive judgements through the calibration of comprehension and the calibration of performance paradigms. In the present study, the measures used to assess calibration of comprehension involve three metacognitive judgements; JOL (i.e. judgments of understanding and confidence), EOL (i.e. judgments of text-easiness) and preCJ (i.e. pre-test confidence judgements). Calibration of performance was assessed through postCJs (i.e. post-test confidence judgements). The students in this study were not able to calibrate their comprehension at above chance level. According to the "Cognitive Effort Hypothesis" of Maki, et al. (1990), if the to-be-learned text is easy for the reader, calibration of comprehension for the text is low since the readers do not attend much to the details. However, in the case of texts those require higher levels of cognitive processing, calibration of comprehension increases. The difficulty level of the texts used in this study was similar to the difficulty level of the texts used in examinations for candidates of BA programmes. As all the students in this study had succeeded this kind of examination, they may have found the texts easy. In further research, investigating students' metacognitive monitoring judgements while reading L2 texts with various levels of difficulty can shed light on the influence of task difficulty on metacognitive monitoring while reading texts in L2.

Low calibration Gammas in this study is also inconsistent with previous research findings in L1 reading. Previous research in L1 reading showed that the readers could calibrate their comprehension at above chance level although calibration Gammas were very low. This difference may be attributed to the lack of knowledge of discourse organisation, which is very important for L2 readers especially in the case of reading advanced level academic texts. Students may not follow the particular way of development of the text, the new information or the arguments presented in the text because of their inefficient discourse knowledge; but, if they are familiar with most of the vocabulary and grammatical structures, they may think that they understand the text well.

In case of calibration of performance, the students in this study were able to calibrate their performance. This result is consistent with previous research findings in L1 reading. Most studies showed that readers' postdictions are more accurate than their predictions. According to Lin et al. (2001) postdictions are more accurate than predictions, because readers use additional information from performing on a test as feedback to make more precise judgements. However, the calibration Gammas in this study are much higher than those reported in L1 reading research. High performance calibration Gammas in this study may be explained by the difference between L1 and L2 reading. According to Grabe and Stoller (2002), readers are more aware of the processes in L2 reading than in L1 reading since L2 is usually learned with conscious effort whereas L1 is learned spontaneously. This difference in awareness may be the reason for higher Gammas in this study by comparison with the studies in L1 reading. Obviously, further studies are needed to investigate students' metacognitive monitoring in L1 and L2 reading within the same study to detect the differences.

Alternatively, high Gammas for postdiction judgements in this study may be attributed to the cultural differences between western students and Turkish students. In a study by Zabrucky et al. (2008), the researchers investigated the Taiwanese students' calibration of performance in L1 reading. The results of this study were rather different from the ones conducted with western students since Taiwanese students' calibration Gammas were much higher. The performance calibration Gammas were found to be higher than .50. The researchers asserted that this difference might be due to cultural differences between Taiwanese and western students. Most of the studies on the accuracy of metacognitive monitoring judgements were conducted with participants from western culture. Little is known about the nature of metacognitive monitoring judgements of people from a cultural background other than that of the western world. Cross-cultural studies would be useful to investigate whether the judgements of metacognitive monitoring are influenced by cultural tendencies.

The finding that there is no significant correlation between students' calibrations and their GCP is consistent with much research conducted in L1 reading (e.g. Dunlosky & Connor, 1997; Dunlosky & Hertzog, 1997; Lin,

et al., 2001). This finding is quite acceptable since, as articulated by Metcalfe (2009) and Thiede and Dunlosky (1994), accuracy of metacognitive judgements is necessary but not sufficient for successful learning if the learner cannot convert these judgements into appropriate strategies for study. Nevertheless, there are studies that found evidence for a significant relationship between students' calibrations and their GCP in L1 reading (e.g. Lin & Zabrucky, 1998; Zabrucky et al., 2008). These conflicting results from several studies indicate that more studies should be conducted to investigate the nature of this relationship.

• • •

Seda Sarac is currently a research assistant in the Faculty of Education at Yildiz Technical University. She has BA in Teaching English as a Foreign Language, MA in Curriculum Development and Instruction. She is, now, preparing her PhD dissertation on metacognition and intelligence. Her research interests are metacognition, self-directed learning, self-regulated learning, language syllabus design, and learner autonomy. ssarac@yildiz.edu.tr

Betul Tarhan is currently a research assistant in the Faculty of Education at Yildiz Technical University. She has BA in Teaching English as a Foreign Language, MA in Teaching Turkish as a Foreign Language. She obtained her PhD in Istanbul University on Teaching English as a Foreign Language. Her research interests are self-directed learning, brain compatible learning and learner autonomy. betarhan@yildiz.edu.tr

References

Baker, L., & Brown, A. (1984). Metacognitive skills and reading. In D. Pearson (Ed.), *Handbook of Reading Research*. (ss. 353-394). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Begg, I. M., Martin, M. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology,* 4, 195-218.

Block, E. L. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly, 26,* 319-343.

Casanave, C. (1988). Comprehension monitoring in ESL reading: A neglected essential. *TESOL Quarterly, 22,* 283□302.

Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development, 53*, 11-28.

Connor, L. T., Dunlosky, J.,& Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, *12*,50-71.

Cromley, J. G. (2005). Metacognition, cognitive strategy instruction, and reading in adult literacy. *Review of Adult Learning and Literacy, 5*, 187-204.

Dunlosky, J., & Lipko, A.R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228-232.

Dunlosky, J., & Rawson, K. A. (2005). Why Does Rereading Improve Metacomprehension Accuracy? Evaluating the Levels-of-Disruption Hypothesis for the Rereading Effect. *Discourse Processes, 40,* 37-56.

Dunlosky, J.,& Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*, *25*,691-700

Epstein, W., Glenberg, A. M, & Bradley, M. (1984). Co-activation and comprehension: Contribution of text variables to the illusion of knowing. *Memory & Cognition,12,* 355-360.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34(10), 906–911.

Glenberg, A. M., & Epstein, W. (*1985). Calibration of comprehension. Journal of Experimental Psychology. Learning, Memory, and Cognition, 11(4), 702–718.*

Glenberg, A.M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory &Cognition,* **15**, 84-93.

Glenberg, A.M., Sanocki, T., Epstein, W., & Morris, C. (l987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116*, 119-136.

Grabe, W., & Stoller, F. L. (2002). Teaching and researching reading. New York: Longman.

Kimmel, S., & MacGinitie, W.H. (1984). Identifying children who use a perseverative text processing strategy. *Reading Research Quarterly, 19,* 162-172.

Kolic-Vehovec, S., & Bajsanski, I. (2006). Metacognitive strategies and reading comprehension in elementary-school students. *European Journal of Psychology of Education,* 21, 439-451.

Lin, L., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction, *Contemporary Educational Psychology, 23,* 345–391.

Lin, L., Moore, D., & Zabrucky, K. M. (2001). An assessment of students' calibration of comprehension and calibration of performance using multiple measures. *Reading Psychology*, 22, 111–128.

Lin, L., Zabrucky, K. M., & Moore, D. (2002). Effects of text difficulty and adults' age on relative calibration of comprehension. *American Journal of Psychology,*15(2),187-198.

Maki, R. H. (1995). Accuracy of metacomprehension judgments for questions of varying importance levels. *American Journal of Psychology, 108*(3), 327-344.

Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. Psychonomic Bulletin and Review,1, 126–138.

Markman, E. M. (1977). Realizing that you don't understand: A Preliminary investigation. *Child Development*, 48, 989-992.

Metcalfe, J. (2009). Metacognitive judgements and control of study. *Current Directions in Psychological Science*, 18(3), 159-163.

Miesner, M. T., & Maki, R. H. (2007). The role of test anxiety in absolute and relative metacomprehension accuracy. European Journal of Cognitive Psychology, 19, 650–670.

Morrison, L. (2004). Comprehension monitoring in first and second language reading. *The Canadian Modern Language Review*, 61(1), 77-106.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance and monitoring accuracy. *Metacognition and Learning, 2,* 159-179.

Nietfeld, J. L., Cao, L., & Osborne, J.W. (2005). Metacognitive monitoring accuracy and student performance in the classroom. *Journal of Experimental Education, 74*(1), 7-28.

Nietfeld, J. L., Enders, C. K., & Schraw, G. (2006). A Monte Carlo comparison of two measures of monitoring accuracy. *Educational and Psychological Measurement, 66,* 258-271.

Maki, R. H., J. M. Foley, W. K. Kajer, R. C. Thompson, & M.G. Willert. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 609-616.

Otero, J. C., Campanario, J.M., & Hopkins, K. D. (1992). The relationship between academic achievement and metacognitive comprehension monitoring ability of

Spanish secondary school students**.** *Educational and Psychological Measurement*, 52, pp 419-430.

Pieschl, S. (2009). Metacognitive calibration—An extended conceptualization and potential applications. *Metacognition and Learning, 4, 3-31.*

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Practice*, 41, 219-225.

Pressley M, & Schneider W. 1997. *Introduction to Memory Development During Childhood and Adolescence.* Lawrence Erlbaum Associates: Mahwah, NJ.

Schraw, G. (2002). Promoting general metacognitive awareness. In H. J. Hartman (Ed.). *Metacognition in Learning and Instruction: Theory, Research and Practice.* (ss. 3–16). The Netherlands: Kluwer Academic Publishing.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review,* 7, 351–371.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66-73.

Thiede, K.W., & Dunlosky, J. (1994). Delaying students' metacognitive monitoring improves their accuracy in predicting the recognition performance. *Journal of Educational Psychology*, 86, 290-302.

Veenman, M. V. J. (2005). The assessment of metacognitive skills: what can be learned from multi-method designs? In: B. Moschner, C. Artelt (Eds.) *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (75-97). Berlin: Waxmann.

Veenman, M. V. J., & Elshout, J. J. (1995). Differential effects of instructional support on learning in simulation environments. *Instructional Science, 22,*363-383.

Weaver, C. A. III, & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition, 23*(1), 12–22.

Weaver, C. A. III, & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition,*23, 12-22.

Wiley, J., Griffin, T., & Thiede, K.W. (2005). *Putting the Comprehension in Metacomprehension.* Journal of General Psychology, 132, 408-428.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology, 132(4), 408–428.*

Zabrucky, K, Agler, L. & Moore, D. (2005). Metacognition in Taiwan: Students' calibration of comprehension and performance. *International Journal of Psychology*, 44 (4), 305-312.

Zabrucky, K., & Moore, D. (1994). Contributions of working memory and evaluation and regulation of understanding to adults' recall of texts. *Journal of Gerontology: Psychological Sciences,* 49, 201-212.

Zhao, Q., & Linderholm, T. (2008). Adult metacomprehension: Judgment processes and accuracy constraints. *Educational Psychology Review,* 20, 191–206.