Comparison Between Students and Teachers in Judgment Accuracy and Cue-Utilization: An Explorative Approach.

A.A. Novakovic (3935620)

Bachelor thesis Utrecht University

BA Educational Sciences 2015/2016

Supervisor: Janneke van de Pol

Abstract

This study takes an explorative approach to comparing students' and teachers' judgment accuracy and their corresponding use of cues. The first task students had to perform, consisted of reading six explanatory texts. The sample of students was divided into two conditions; the no diagram and the diagram completion task. In the diagram completion task, students were provided with a pre-structured diagram for each text. The students could score on three cue-utilizations; omission error, commission error, and correct. The no diagram condition got a filler task, consisting of a puzzle. Both students and teachers were asked to provide judgments of comprehension for each text the student has read. The title of the text was presented, with the question of how many questions of the text they expect to answer correct on the test, which they got afterwards. An independent samples t-test compared students and teachers on their judgment accuracy. Differences between students and teachers regarding judgment accuracy was measured by a MANOVA, explaining the differences in cue-utilization. The ANCOVA moderation analyses showed the effect of cue-utilization on the relation between judgment accuracy and students and teachers. Lastly, a t-test was performed to compare the diagram completion task with the no diagram condition.

The differences between students and teachers regarding judgment accuracy, measured by a t-test, was found significant. Teachers scored higher on judgment accuracy than the students. The MANOVA showed a significant difference for cue-utilization count omission and commission. Correct was not significant and is excluded from further analyses. The ANCOVA moderation analyses showed that the cues omission and commission had a positive effect on the relation between judgment accuracy and students and teachers. Differences in the scores of the diagram completion task and the no diagram condition, measured by a t-test, is found significant for the diagram completion task.

Word count: 6095

Index

In the last decades, the focus of education has shifted from purely transferring knowledge, to teaching students to guide their own learning process (Delfino & Persico, 2009; Thomas & Brown, 2011). This view emphasizes on metacognitive skills. Metacognitive skills concern the procedural knowledge and executive skills that are required for self-regulation and monitoring of one's learning activities (Brown & DeLoache, 1978; Flavell, 1992). Currently, new forms of education are rising which presume metacognitive and self-regulated learning skills of students.

However, there has not been much comparative research about which actor is more skilled to give direction to the learning process: the student or the teacher? An essential part of this skill consists of making accurate judgments about learning, which is the main focus of this study (Nelson & Narens, 1990; Schneider, 2008).

Research on making judgments builds on the well-acknowledged cue-utilization framework of Koriat (1997), which states that to judge their learning, people use cues that are accessed prior to making a judgment. Examples of such cues are the perceived relative difficulty of the study items or the type of test expected (Thiede, Griffin, Wiley, & Anderson, 2010). Because cues are used to make a judgment, the judgment accuracy will be determined by how well those cues predict test performance, i.e., cue-diagnosticity (Brunswik, 1956; Koriat, 1997). When cues are used that are more diagnostic of subsequent test performance, judgment accuracy will improve (Thiede et al., 2010).

The primary goal of this study is to compare students' and teachers' judgment accuracy and their corresponding use of cues. First, a detailed overview of the process and importance of making accurate judgments for both students and teachers will be presented. Subsequently, the necessity of this comparison is described, as well as theoretical predictions for this study.

Students' judgments

Importance of students' judgment accuracy. Previous researchers used various definitions for more or less the same concepts. Therefore, students' judgments of comprehension (SJOC) are defined as judgments made by students regarding their ability of recalling and applying information from an explanatory text on a subsequent test (Koriat, 1997). Those judgments are said to be accurate, when they are consistent with objective assessment of the same skills (Ready & Wright, 2011). Students who can accurately judge their level of understanding are able to learn more from textual information (Dunlosky & Rawson, 2012; Thiede, Anderson, & Therriault, 2003). In particular, if students can judge what material they have understood well and what they have not, they can focus their attention just on the not-understood information (e.g., Dunlosky, Hertzog, Kennedy, & Thiede, 2005). Therefore, students' judgment accuracy is critical for continued strategy use, making study decisions and consequently learning efforts (Metcalfe & Finn, 2008; Thiede et al., 2003). If students' judgment accuracy is poor, they will not be able to use their judgments to appropriately guide their own learning.

Improving students' judgment accuracy when learning from text. Unfortunately, students' monitoring of text comprehension is often inaccurate (De Bruin, Thiede, Camp, & Redford, 2011; Dunlosky & Lipko, 2007). Multiple studies have shown that to improve accuracy when monitoring understanding of a text, learners need to base their judgments on cues that arise from processing information about the gist of a text (e.g., Thiede et al., 2010; Rawson, Dunlosky, & Thiede, 2000). With respect to understanding of expository texts, this gist comprehension mainly depends on a reader's ability to connect and understand the cause-and-effect relations in a text (Graesser et al., 1994).

Delayed diagram completion task. An intervention that helps students focus on causal relationships is the diagram completion task, as used by Van Loon, De Bruin, Van Gog, Van Merriënboer and Dunlosky (2014). The results of their study show that this task

provides learners with cues that indicate whether they have understood the cause-and-effect relations within the text. Because those cues are diagnostic of subsequent performance, this task improves the accuracy of their comprehension judgments (Van Loon et al., 2014).

Delayed diagram completion supported higher judgment accuracy in comparison to immediate diagram completion (Van Loon et al., 2014). This is in line with prior research, suggesting that cues produced by a task vary in diagnosticity. There is a notable difference between the level of mental representation involved in completing the diagram task immediately versus completing the diagram task with a delay (Thiede, Dunlosky, Griffin, & Wiley, 2005). For complete text comprehension, learners must go beyond text base processing of factual information and establish a coherent mental representation of the gist of the text, i.e. a situation model (Kintsch, 1998). Through delay, the diagram completion task helps readers focus on the quality of their situation model, which yields diagnostic cues (Van Loon et al., 2014). However, these findings have not been verified yet and replication studies are valuable for the reliability of results (John, Loewenstein, & Prelec, 2012; Lakens, Haans, & Koole, 2012). A secondary aim of this study is therefore to verify whether the diagram completion task indeed supports higher judgment accuracy.

Cue-utilization. Based on the diagram completion task, Van Loon et al. (2013), assume the possible presence of four diagnostic cues: the extent to which correct causal relationships were provided in the diagram (correct), the extent to which provided answers were not based on the text (commission error), the extent to which no response was given (omission error), and lastly, the extent to which factual information was provided instead of a correct step in the causal chain (factual information). Building upon their study, three cues are taken into account in this study. The last cue has not been taking into account for this study and the reason is because understanding causal relations requires more insight than

learning facts. Students learn lists of facts at school, but acquire minimal understanding of important causal relations in a text (Wolfe & Goldman, 2005).

Teachers' judgments

Importance of teachers' judgment accuracy. Teachers also make ongoing judgments about students' understanding (e.g., Alvidrez & Weinstein, 1999). In this study, teachers' judgments of comprehension (TJOC) are defined as judgments made by a teacher regarding a students' ability of recalling and applying information from a text on a subsequent test. The ability to accurately assess students' performance is considered to be an important aspect of teachers' professional competence, because these judgments guide instructional decisions that may affect students' performance (Ready & Wright, 2011; Südkamp, Kaiser, & Möller, 2012). Specifically, more accurate judgments could lead to better differentiation of instruction, which produces greater gains in students' learning (Thiede et al., 2015). Moreover, judgments influence teachers' expectations about students' abilities (e.g., Brophy & Good, 1986), students' academic self-concept (e.g., Möller, Polmann, Köller, & Marsh, 2009), and it identifies struggling students (e.g., Bailey & Drummond, 2006).

Teachers' judgment accuracy: State of the Art. A recent meta-analysis of 75 articles about teachers' judgment accuracy, conducted by Südkamp, Kaiser, and Möller (2012), yielded a median correlation of .53 of relative judgment accuracy. Their results show that teachers' judgments are far from perfect and that there is plenty of room for improvement. Remarkably, a lot of variation in teachers' judgment accuracy could not be explained, suggesting that teachers vary widely in their judgment accuracy. Understanding these different levels of accuracy is complicated by the fact that researchers have used a variety of approaches to compute the correlation between predicted and actual performance (Thiede et al., 2015). As in prior research, the focus of this study is placed on relative

accuracy, which is the degree to which predictions discriminate between the different levels of performance on the test for one text relative to another. Thiede et al. (2015) recommends the use of intra-individual gamma correlation, because this provides a measure of an individual teachers' ability to differentiate levels of learning among the students in his or her own classroom (like Helmke & Schrader, 1987 did).

Cue-utilization. There are a variety of cues available for teachers to judge students' comprehension (Thiede et al., 2015). This study builds upon the cues presented by Van Loon et al. (2014), based on the diagram response categories as described above (correctly stated causal relations, commission errors and omissions). Van Loon et al. (2014) is the only study that has been conducted from this particular subject. That is why the same cues are expected to be used by teachers to make judgments about the comprehension of their student.

The current study

In the current study, judgment accuracy and cue-utilization of both students and teachers is examined. As described above, different studies have used a variety of approaches to measure students' and teachers' judgment accuracy, which makes comparison challenging. A study is needed which measures those aspects in a similar way, so that results can be compared. The focus of this study is placed on relative accuracy and the use of intra-individual gamma correlations for the students and teachers and thereby makes comparing possible.

The main research questions are: 1. To what extent are there differences between students and teachers regarding judgment accuracy? 2. Can potential differences in judgment accuracy between teachers and students be explained by differences in cue-utilization? This study is the first to make this comparison and will therefore take an explorative approach. Previous studies have shown various results in accuracy (De Bruin, Thiede, Camp, &

Redford, 2011; Dunlosky & Lipko, 2007; Südkamp et al., 2012; Thiede, 2015; Van Loon; 2014). However, there has not been enough evidence to give direction to the hypotheses. Therefore, we hypothesize that no significant differences are expected between teachers' and students' cue-utilization and no significant differences are expected between teachers' and students' judgment accuracy.

Because the design of this study relies much on that on Van Loon et al. (2014), a third research question concern replicating their findings: 3. To what extent does the delayed diagram-completion task lead to more accurate students' judgments in comparison to omission of this task? Based on the results of Van Loon et al. (2013), judgment accuracy is expected to be higher when students complete the delayed diagram completion task, relative to when they do not.

Method

Participants

Both students and teachers have participated in this study. By means of a convenience sample, fifteen teachers of various secondary schools across the Netherlands were recruited. The fifteen participating teachers were between 24 and 58 years (M = 40.40, SD = 10.90), of which 66.7 percent was female and 33.3 percent male. Teachers from various subjects participated; the inclusion criterion was that reading and studying explanatory texts was an essential part of the subjects' curriculum. The average years of teaching experience of the teachers was 14.88 (SD = 9.36) and ranged from 2 to 35 years with an average of 8.9 months of exposure to the students they had to judge in this study.

Each teacher was asked to select one of their classes to participate in this study through a convenience sample. 181 high school students participated from six Dutch secondary schools, of which 62.4 percent were female and 37.6 percent male. All students

were between 11 and 16 years old (M = 14.58, SD = 0.65) and they were all third year students of secondary education, following one of the two educational programs that lead to higher education. 75 students followed the pre-university program (VWO; highest level of secondary education), and 106 students followed the higher general secondary education program (HAVO; middle level of secondary education).

Materials

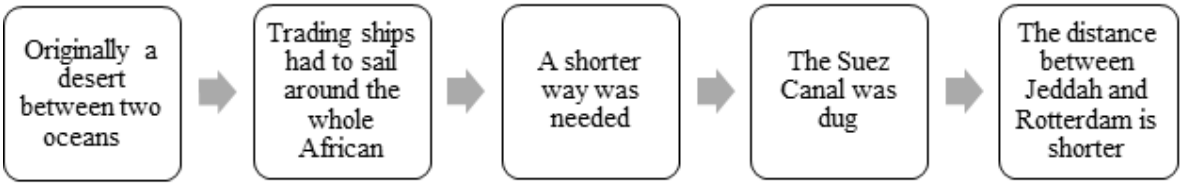
Table 1 is a schematic overview presented of the various phases of this study. The materials that were presented during each phase will be described below.

Table 1
Schematic overview of the various phases of this study

Students			
Condition	**Diagram completion**	Practice session	**No diagram completion**
	Read text 1		Read text 1
	Read text 2		Read text 2
	…		…
	Read text 6		*Read text 6*
	Diagram text 1		*Filler task*
	Diagram text 2		
	…	*Students' Judgments of Comprehension*	
	Diagram text 6	*Test*	
Teachers			
		Practice session student materials	
		Practice session Teachers' Judgments of Comprehension	
		Teachers' Judgments of Comprehension	

Text study. Based on Van Loon et al. (2014), six explanatory texts were selected for the experiment in which both causal and factual relations were presented (see Appendix A for an example of a text). Causal relationships could occur in both serial and simultaneous formats and each texts contained five (elements of) causal relations. The topics of the texts were “Sink”, “Botox”, “The Suez Canal”, “Music makes smart”, “Money does not make happy” and “Renovation”. There were also six versions of the text booklets, and the difference between them is the text sequence.

Diagram completion task. In this task, students were provided with a pre-structured diagram for each text. All diagrams contained five textboxes, of which one was already filled in, representing either serial or simultaneous causal relations (see Figure 1 for a completely filled-in example diagram). The students were asked to fill in the four empty textboxes.



Figuur 1. A correctly completed diagram for the text “Suez Canal”

Students’ responses in the boxes of the diagrams were classified into the three following categories. A response was scored as correct when answers literally showed the causal relations or showed gist understanding of the text. Commission error was scored when incorrect causal relations were established or vague answers were given. When students didn’t fill in a box, this was scored as omission error. With respect to the scoring of questions about causal relationships, a response model was compiled including the correct causal relationships from each text. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among raters, which was a substantial agreement, κ = .75, $p < .05$. For each question about the causal relationships from one text, two scores were

computed. Firstly, a score that indicates the number of correctly stated causal relationships. Secondly, a score indicating the number of causal relationships that were not mentioned. Both scores range from 0 to 4, since students had to identify four causal relations from each text.

Students' Judgments of Comprehension (SJOC). Students were asked to provide two SJOC's for each text, one about the causal relations and one about the factual information. The title of the text students had just read was presented to them, accompanied by the following questions:

1. How many questions concerning the causal relationships of this text do you expect to answer correctly on the test?

The response scale for this question ranged from 0 to 4.

2. How many questions concerning the facts of this text do you expect to answer correctly on the test?

The response scale for this question ranged from 0 to 5.

Teachers' Judgments of Comprehension (TJOC). The teachers also provided two TJOC's for each text the student has read. They were presented similar questions, but then about the students' understanding of causal relationships and facts from the texts, and the same answer scales as the students.

Test. Students were tested for their understanding of causal relationships and their remembrance of facts from each text with a test. This test included for each text five questions about facts and one question about the causal relationships of the texts. In the latter question, students were asked to identify four causal relations from the corresponding text.

The students' responses of the factual information questions were scored in two ways. The number of boxes that were filled in with correct relations are scored and the boxes that were not filled in were scored. With respect to the scoring of questions about causal relationships, a response model was compiled including the correct causal relationships from each text. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among raters that scored test performance, which was an almost perfect agreement; $\kappa = .93$, $p < .05$.

Design and procedure.

A between-subjects design was used to compare students and teachers on their judgment accuracy, which is the first research question, and cue-utilization, that implies the second research question. To confirm whether or not the delayed diagram completion task was indeed valuable for making accurate judgments, an experimental between-subjects design was used. Students were randomly assigned to one of the two conditions: diagram completion ($N = 151$) and no diagram completion ($N = 30$).

Table 1 depicts the procedure of this study for both students and teachers. Students completed the tasks in one session that lasted for approximately one hour, which took place in their own classroom. In the practice session, students were first instructed about the type of texts, the distinction between causal and factual information, students' judgment on performance, the diagram completion task and the test format. Students were instructed that they would study six texts for a later performance test with questions on factual information and causal relations, and that they would be asked to judge their comprehension of these texts by predicting future test performance. In the delayed diagram completion group, students first read all six texts, and then started with the diagram completion task of each text. The students in the no-diagram group completed a filler task after reading all six texts, instead of completing diagrams, and finished the experiment by predicting their future test

performance. The filler tasks consisted of six duplicated pictures that are related to the six texts they read, where the students had to find the four differences between the two pictures. The no-diagram group was taken into account so that comparison between the diagram and no diagram completion groups is possible. This comparison is related to the third research question.

Teachers also started with a practice session about the students' materials and the procedure of the task. Following this, teachers practiced with predicting students' performances by estimating the performances of two random students. The teachers were provided with the students' completed diagram (of students within the diagram completion condition) and were instructed to base their predictions of student performance on the diagram. After this practice session, all teachers provided judgments of comprehension for each text for fourteen of their students. The practice session took place in the classroom and lasted about fifteen minutes.

Before the experiment, all students were asked to join voluntarily through a consent form for both students and their parents. Teachers and schools also participated on a voluntary base. All data collected in the experiment was processed anonymously, and interpreted with care and precision. Results are not to be traced back to individual persons.

Analyses. As prior research recommends, relative judgment accuracy was measured for students and teachers by gamma correlations (Van Loon et al., 2014; Thiede et al., 2003). Gamma indicates the strength of the relation between students' and teachers' judgments and their actual test performance and ranges from -1 (indicating a perfect negative association) to +1 (indicating a perfect positive association). A gamma of zero indicates that there is no relation between judgments and actual performance (Van Loon et al., 2014). Cue-utilization is scored in the same way (ranges from -1 to +1) as judgment accuracy and is operationalized as the gamma correlation between the judgments of student learning and the cues (responses) in the diagram (Van Loon et al., 2014).

Below, for each research question the intended analyses are described.

1. To what extent are there differences between students and teachers regarding judgment accuracy? Judgment accuracy is operationalized as the Goodman and Kruskal's gamma correlation between the test score and the altitude of the judgment. To examine this question, an independent samples t-test will be performed with judgment accuracy as the dependent variable teachers and students as independent variables.

2. Can potential differences in judgment accuracy between teachers and students be explained by differences in cue-utilization? Cue-utilization is operationalized as the Goodman and Kruskal's gamma correlation. This research questions consists of two parts. First, differences in cue-utilization need to be examined. This will be done with a MANOVA, comparing both students and teachers on the three dependent variables of cue-utilization; omission error, commission error and correct. When variables of cue-utilization are not significant, they will not be taken into account for the second part of this research question. Second, to examine whether the relation between judgment accuracy and student or teachers is influenced by cue-utilization, an ANCOVA moderation analysis will be performed with the variables of cue-utilization that are significant. Both judgment accuracy and cue-utilization are operationalized as the Goodman and Kruskal's gamma correlation. However, when the difference in judgment accuracy is not significant, only the cue-utilization will be analyzed.

3. To what extent does the delayed diagram completion task lead to more accurate students' judgments in comparison to the no diagram condition? To answer this question, an independent samples t-test will be performed between students in the two conditions. The conditions of the delayed diagram-completion task versus no diagram completion are the independent variable and the students' judgment accuracy, operationalized as the Goodman

and Kruskal's gamma correlation, is the dependent variable. All analysis will be performed with IBM SPSS Statistics 24, using a p-value of .05.

Results

Judgment accuracy. An independent-samples t-test was conducted to compare TJOC with SJOC (see table 2). Levene's test for equality of variances was significant (F = 15.83, $p < .05$), indicating that the assumption of homogeneity of variances has been violated and equal variances between the groups cannot be assumed. A significant difference was found in the scores for the TJOC (M = .17, SD = .68) and SJOC (M = .00, SD = .61); $t(1560.30) = 5.173$, $p < .05$, suggesting TJOC was higher than SJOC. TJOC indicates a positive association and is more accurate, where SJOC indicates no relation between judgment and actual performance. These results reject the hypothesis, indicating that no differences between students and teachers were expected.

Table 2

Independent samples test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig.	Mean Differen-ce	Std. Error Differen-ce	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	15,829	,000	5,150	1561	,000	,167973	,032619	,103992	,231954
Equal variances not assumed			5,173	1560,30	,000	,167973	,032470	,104284	,231662

Cue-utilization. Differences in cue-utilization between students and teachers were examined on three dependent variables of cue-utilization; omission error, commission error and correct, and measured with a MANOVA. A significant difference has been found in cue-utilization between students and teachers, $F(3, 542) = 57.76$, $p = < .05$ (see table 3). This rejects the hypothesis that no significant differences would be found in cue-utilization between teachers and students. Variable omission error indicates a negative association for both teachers ($M = -.29$, $SD = .71$) and students ($M = -.29$, $SD = .65$) between judgments of students' learning and cues in the diagram. Commission error shows a positive association for both teachers ($M = .39$, $SD = .64$) and students ($M = .25$, $SD = .57$) between judgments of students' learning and cues in the diagram, whereas the variable correct indicates a negative association for teachers ($M = -.19$, $SD = .59$) and no relation for students ($M = -.01$, $SD = .62$) between judgments of student learning and cues in the diagram.

Table 3

Differences in cue-utilization

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	,242	57,763	3,000	542,000	,000
	Wilks' Lambda	,758	57,763	3,000	542,000	,000
	Hotelling's Trace	,320	57,763	3,000	542,000	,000
	Roy's Largest Root	,320	57,763	3,000	542,000	,000
Person	Pillai's Trace	,026	4,871	3,000	542,000	,002
	Wilks' Lambda	,974	4,871	3,000	542,000	,002
	Hotelling's Trace	,027	4,871	3,000	542,000	,002
	Roy's Largest Root	,027	4,871	3,000	542,000	,002

A significant difference has been found for cue-utilization count omission, $F(1, 544) = 8.58$, $p = < .05$ and cue-utilization count commission, $F(1, 544) = 4.34$, $p = < .05$. Cue-utilization count correct was not significant, $F(1, 544) = 1.34$, $p = > .05$ and will be excluded from further analyses. These results are represented in table 4.

Table 4

Differences between cue-utilization

	F	df1	df2	Sig.
Omission	8,580	1	544	,004
Commission	4,342	1	544	,038
Correct	1,336	1	544	,248

The ANCOVA moderation analyses is performed with the variables of cue-utilization that are significant and showed the influences of cue-utilization on the relationship between judgment accuracy and students and teachers. The covariate, cue-utilization count omission, was significantly related to judgment accuracy, $F(1,561) = 6.35$, $p < .05$, $r = .11$, which indicates a small effect (see table 5).

Table 5

Covariate cue-utilization count omission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3,423	2	1,711	4,578	,011	,016
Intercept	14,824	1	14,824	39,659	,000	,066
Commission	2,375	1	2,375	6,353	,012	,011
Person	1,293	1	1,293	3,459	,063	,006
Error	209,693	561	,374			
Total	242,198	564				
Corrected Total	213,116	563				

The covariate, cue-utilization count commission, was significantly related to judgment accuracy, $F(1,513) = 52.83$, $p < .05$, $r = .30$, which indicates a medium effect (table 6).

Table 6

Covariate cue-utilization count commission

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	17,991[a]	2	8,995	28,699	,000	,101
Intercept	9,414	1	9,414	30,034	,000	,055
Omission	16,560	1	16,560	52,834	,000	,093
Person	1,352	1	1,352	4,314	,038	,008
Error	160,792	513	,313			
Total	205,409	516				
Corrected Total	178,782	515				

Diagram completion task. An independent t-test was conducted to compare the delayed diagram completion task with the no diagram condition to examine if the diagram completion task leads to a more accurate SJOC. Levene's test for equality of variances was not significant ($F = .2.76$, $p > .05$), indicating that the assumption of homogeneity of variances has not been violated and equal variances between the groups can be assumed (see table 7).

A significant difference in the scores is found for the diagram completion task ($M = -.02$, $SD = .59$) and no diagram condition ($M = .13$, $SD = 0.84$); t (746)=2.65, $p = < .05$. The hypothesis stated that students who complete the delayed diagram task have higher judgment accuracy than those who do not is rejected. The no diagram condition scored higher on SJOC than the delayed diagram completion task. Compared to the no diagram condition, the diagram completion task did not contribute to a higher level of students' judgment accuracy and did not have a positive effect on SJOC.

Table 7

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	2,763	,097	2,653	746	,008	,152619	,057523	,039693	,265546
Equal variances not assumed			2,465	183,712	,015	,152619	,061924	,030445	,274794

Conclusion

Results suggest that there is a difference between students and teachers regarding judgment accuracy. TJOC indicates a positive association and is more accurate whereas SJOC indicates no relation between judgment and actual performance. These results reject the hypothesis that signifies no differences between student and teachers were expected.

Differences in judgment accuracy between teachers and students can be explained by differences in cue-utilization. Both students and teachers were first compared on the three variables of cue-utilization; omission error, commission error and correct. Omission and commission error are found significant and the variable correct is not significant, meaning that the variable correct was excluded from further analyses. Further, the variables omission

and commission of cue-utilization has a moderating effect on the relation between judgment accuracy and students or teachers. Cue-utilization omission is significantly related to judgment accuracy with a small effect and cue-utilization commission is significantly related to judgment accuracy with a medium effect. These results reject the hypothesis, saying that no significant differences would be found in cue-utilization between teachers and students.

Delayed diagram completion task does not lead to more accurate students' judgments in comparison with the no-diagram condition. The results show a significant difference in the scores, with the no diagram condition scoring higher on SJOC than the diagram completion task. When compared to the no diagram condition, the diagram completion task does not contribute to a higher level of students' judgment accuracy and does not have a positive effect on SJOC. The hypothesis, stated that students who complete the delayed diagram task have higher judgment accuracy than those who do not, is rejected.

Discussion

The aim of this study is to explore the fields of judgment accuracy and cue-utilization and examine possible differences in these variables between students and teachers. Prior research indicates that students' judgment accuracy for learning of causal relations from texts would improve if the students completed diagrams prior to making judgments about their test performance, in comparison to students who did not use diagrams (Van Loon et al., 2014). Another aim of this study is to demonstrate this improvement.

This study was the first to compare students' and teachers' judgment accuracy and therefore took an explorative approach. Teachers' and students' judgment accuracy was separately measured in prior research, but was never compared to each other.

Previous research mostly found that students' judgment of text comprehension is often inaccurate (De Bruin et al., 2011; Dunlosky & Lipko, 2007). A significant difference is

found between students and teachers, indicating that teachers scored relatively low on judgment accuracy and no relation is found for students between their judgment and actual performance. Teachers scoring relatively low on judgment accuracy indicates that their ability to accurately assess students' performance is low, meaning that they possibly have trouble with guiding instructional decisions that may negatively affect students' performance (Ready & Wright, 2011; Südkamp, Kaiser, & Möller, 2012). With these findings, teachers' expectation about the ability of students to perform (e.g., Brophy & Good, 1986) and identify students that are struggling (e.g., Bailey & Drummond, 2006) can be inaccurate as well.

No relation is found for students between their judgment and actual performance, indicating that their ability of recalling and applying information from an explanatory text on a subsequent test is poor (Koriat, 1997). Poor judgment leads to not knowing what material students have understood and what they have not, meaning that they cannot focus on the information they did not understand (e.g., Dunlosky, Hertzog, Kennedy, & Thiede, 2005). They will not be able to use their judgments to appropriately guide their own learning. Students' accurate judgment is needed for continued strategy use, making study decisions and consequently learning efforts (Metcalfe & Finn, 2008; Thiede et al., 2003).

The differences between students and teachers regarding judgment accuracy, can be explained by differences in cue-utilization. Results of cue-utilization imply a significant difference in cue-utilization between teachers and students, whereas the teacher had a higher score. These results reject the hypothesis, saying that no significant differences would be found in cue-utilization between teachers and students. Comparison between students and teachers is made with the three variables of cue-utilization, resulting in a significant effect for omission and commission. The variable correct is not significant, and therefore excluded from further analyses. Both variables omission and commission of cue-utilization have a moderating effect on the relation between judgment accuracy and students or teachers.

Van Loon et al. (2013), is the only study that has been conducted from this particular study and assume the possible presence of four diagnostic cues. The four cues consist of the three cues used in this study, and the fourth is factual information. The variable factual information has not been included in this study, considering understanding that causal relations requires more insight than learning facts (Wolfe & Goldman, 2005). Both students and teachers scored low on the causal relations, and therefore it would be a good option to add the fourth variable of factual information and compare factual information with causal relations.

Based on the results of Van Loon et al. (2013), judgment accuracy is expected to be higher when students complete the delayed diagram completion task, relative to when they do not. Results of this study show a significant difference in the scores, with the no diagram condition scoring higher on SJOC than the diagram completion task. When compared to the no diagram condition, the diagram completion task does not contribute to a higher level of students' judgment accuracy and does not have a positive effect on SJOC. Delayed diagram completion task did not lead to a more accurate judgment of students in comparison with the no-diagram condition, as was concluded by Van Loon et al. (2014). Therefore, the hypothesis is being rejected, stating that students who complete the delayed diagram task have higher judgment accuracy than those who do not.

The findings of prior research cannot be applied to this study. Van Loon, De Bruin, Van Gog, Van Merriënboer and Dunlosky (2014), implied that the diagram completion task is an intervention that helps students focus on causal relations. Van Loon et al., (2014) also implied that through delay, the diagram completion task will help readers focus on the quality of their situation model, which yields diagnostic cues (Van Loon et al., 2014).

A possible explanation for why the delayed diagram completion task did not contribute to the judgment accuracy of students and teachers, could be the large amount of missing data and

distribution of the participants among the conditions. The results could have been influenced by the number of participants per condition and the samples could have been too small. Research on the diagram completion task is necessary to confirm the positive influence on judgment accuracy.

Both students and teachers scored low on causal relations. Therefore it would be appealing to take the fourth variable of factual information into further research and compare factual information with causal relations. Students with dyslexia should be taken into account, whereas this could influence the validity of the study.

For further research, it is important to take in account that the groups being compared have an equal amount of participants in all groups. It makes comparing more reliable.

The intern validity of this study is ensured through randomly assigning teachers and students to different conditions to assure that the groups were similar to each other. Procedures are carefully explained with examples and exercises to make sure that students and teachers comprehended how to execute the tasks. The external validity of the results could have been higher, using a more variating group of students. The current study is limited to HAVO and VWO students from the third grade, which implies a very specific group of students. It is recommended that future research broadens this view. Results could be different when the study would include more variety within the samples of students, differing in ages, education levels, and social backgrounds. Almost all students and teachers in this study have a Dutch nationality. Therefore, one should be careful with generalizing these results to other countries.

References

Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of educational psychology*, *91*(4), 731-746.

Bailey, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment*, *11*(3-4), 149-178.

Bruin, A. B. de, Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology*, *109*(3), 294310.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Unity of California Press.

Delfino, M., & Persico, D. (2009). Self-regulated learning: issues and challenges for initial teacher training. *Handbook of Research for New Media Literacy at the K-12 Level: Issues and Challenges*, 839-845.

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. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271-280.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review*, *101*(3), 371-395.

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-

explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*(1), 93-103.

John, L.K., Loewenstein, G., & Prelec, D. (2012). Questionable Research Practices. *Psychological Science, 23(5),* 524-532.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. J*ournal of Experimental Psychology: General, 126*(4), 349-370.

Lakens, D., Haans, A, & Koole, S. (2012). Eén Onderzoek is Geen Onderzoek. *De Psycholoog, 9,* 10-18.

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1084-1097.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research,79*(3), 1129-1167.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.). The psychology of learning and motivation. New York, USA: Academic Press.

Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition, 28*(6), 1004-1010.

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of

young children's cognitive abilities the role of child background and classroom context. *American Educational Research Journal*, *48*(2), 335360.

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, *2*(3), 114-121.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743-762.

Thiede, K. W., & Anderson, M. C. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, *28*(2), 129-160.

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of educational psychology*, *95*(1), 66-73.

Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., & Jesse, D. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education*, *49*, 36-44.

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1267-1280.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47*(4), 331-362.

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring

during and after reading. *Handbook of metacognition in education*, 85-106.

Thomas, D., & Brown, J. S. (2011). *A new culture of learning: Cultivating the imagination for a world of constant change* (Vol. 219). Lexington, KY: CreateSpace.

Van Loon, M. H., de Bruin, A. B., van Gog, T., van Merriënboer, J. J., & Dunlosky, J.(2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta psychologica, 151*, 143-154.

Wolfe, M. B., & Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction*, *23*(4), 467-502.

Appendix A

Text “The Suez Canal”

The Suez Canal, which connects the Indian Ocean and the Mediterranean Sea with each other, is of great importance to the world. Originally, there was no natural water connection between the Atlantic and the Indian Ocean. Between these two seas is a desert. This meant that trading ships that traveled from the harbor city Jeddah in Saudi Arabia to Europe had to make a long journey around the whole African continent. It was therefore decided that a shorter waterway was needed that would connect the two oceans with each other. For this reason, the Suez Canal, which was designed by the Austrian engineer Alois Negrelli, was dug. For years, workers were digging; the canal was finally opened in 1869 for shipping. By the digging of the Suez Canal, the distance from the harbor city of Jeddah to the harbor city of Rotterdam has been reduced by 40%. Through the Suez Canal, the distance between these cities is 6,337 nautical miles, when ships sail around the African continent this distance is 10,743 nautical miles.