Computers & Education 75 (2014) 229–252

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Why fifth- and seventh-graders submit off-task responses to a web-based reading comprehension tutor rather than expected learning responses

Bonnie J.F. Meyer[a,*], Kausalai Kay Wijekumar[b]

[a] Department of Educational Psychology, Counseling, and Special Education, The Pennsylvania State University, 205 Cedar Building, University Park, PA 16802, USA

[b] The Pennsylvania State University, College of Information Sciences and Technology, Penn State Beaver, 212 Ross Administration Building, Monaca, PA 15061, USA

ARTICLE INFO

Article history:
Received 28 September 2013
Received in revised form
28 February 2014
Accepted 28 February 2014
Available online 12 March 2014

Keywords:
Human–computer interface
Intelligent tutoring systems
Pedagogical issues
Teaching/learning strategies

ABSTRACT

Research shows the students improve their reading comprehension with Intelligent Tutoring of the Structure Strategy (ITSS). One problem for ITSS is that some students are producing responses in the on-line instruction that are unrelated to learning and practicing the reading strategy. These types of disengaged responses can be referred to as system active off-task responses ("off-task"). In this study we characterize who produces off-task responses and why. Classification and Regression Trees (C&RT) and logistic regression analyses were used to answer the why question. Variables predicted to relate to gaming included reading strategy and skill variables, motivation, attitude, self-efficacy, and goal orientation variables, demographic variables, and type of computer feedback (simple versus elaborated). C&RT analysis could explain 66% of the variance in off-task responses. Students without off-task responses were higher in motivation to read and worked in ITSS to produce good main ideas. Students with higher off-task responses had low scores on work mastery goals. The highest producers of off-task responses in Grades 5 and 7 (averaging 24 off-task responses over 7 lessons) had low motivation to read and scored over 2 SD below average on recall tasks in ITSS. The logistic regression could explain 42% of the variance in off-task responses. Use of motivational scales prior to starting instruction as well as on-line performance measures could be used to flag students for early intervention to prevent system active off-task responses and increase on-line learning. The C&RT approach may be particularly helpful to designers in making software more appropriate for different types of students.

1. Introduction

The Intelligent Tutoring of the Structure Strategy (ITSS) provides on-line instruction in reading comprehension for upper elementary and middle school students in content area reading. The structure strategy teaches students about five common text structures (e.g., problem-and-solution; comparison) used in expository texts and how to use this knowledge strategically to increase learning, memory, and writing about text ideas. Our ultimate goal for teaching children the structure strategy is that with practice this strategic approach to using text structure will become an automatic skill available for the purpose of close reading of complex expository and persuasive texts. This goal is compatible with Common Core State Standards (2010). These standards list understanding text structures, constructing main ideas, writing summaries, and other skills related to text structure under both English literacy and scientific and technical literacy for upper elementary and middle school students. ITSS was initially designed to provide easily accessible structure strategy instruction via user-friendly web-based training (Meyer & Wijekumar, 2007; see Figs. 1 and 2 for screen shots of ITSS interface).

Research shows that students improve their reading comprehension by using ITSS (Meyer et al., 2010; Meyer, Wijekumar, & Lin, 2011; Wijekumar, Meyer, & Lei, 2012, 2013). However, there are students who do not produce on-task responses and do not work to correct their

* Corresponding author. Tel.: +1 814 234 3944.
E-mail addresses: bjm8@psu.edu (B.J.F. Meyer), kxw190@psu.edu (K.K. Wijekumar).

http://dx.doi.org/10.1016/j.compedu.2014.02.013
0360-1315/© 2014 Elsevier Ltd. All rights reserved.

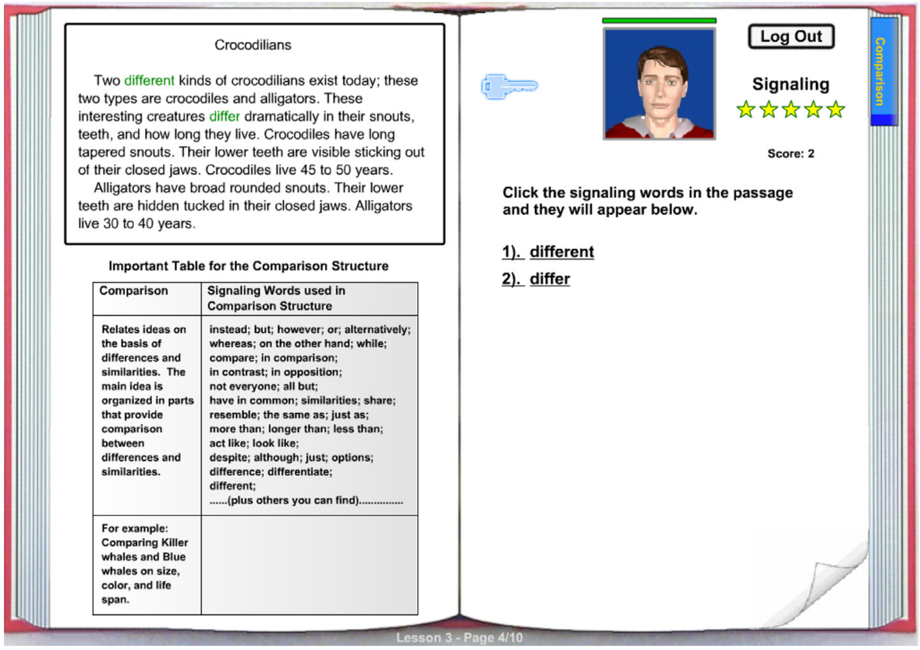

Fig. 1. ITSS interface with instruction about comparison signaling words.

performances via the instruction and feedback provided by ITSS. Students who frequently produce off-task responses while working in the ITSS system miss an opportunity to improve their reading comprehension with expository texts, a critical skill throughout formal schooling and across the lifespan. Such off-task responses are a problem for ITSS and other intelligent tutoring systems in that some students are engaged in avoiding deep thinking about the instruction instead of engaged in learning. Table 1 summaries some of the extant categories used by other researchers to examine disengagement behaviors related to intelligent tutoring systems (ITS). Other categories have been outlined, such as carelessness (Wixon, 2013), but they lack relevance to the problematic responses in the log files from ITSS. The most similar category is listed in the second row of Table 1, responses that are unrelated to the goals of the instruction (Rowe, McQuiggan, Robinson, & Lester, 2009; Wixon, 2013; Wixon, de Baker, Gobert, Ocumpaugh, & Bachmann, 2012). In line with Rowe et al. (2009) and the foundational research by Carroll (1963), we will call these system active, disengaged responses by the term, **off-task responses**. As a step toward reducing such off-task responses, in this study we identify students who produced these off-task responses in a sample of fifth- and seventh-grade students and primarily focus on learner characteristics that may explain why they were engaged in these behaviors rather than learning.

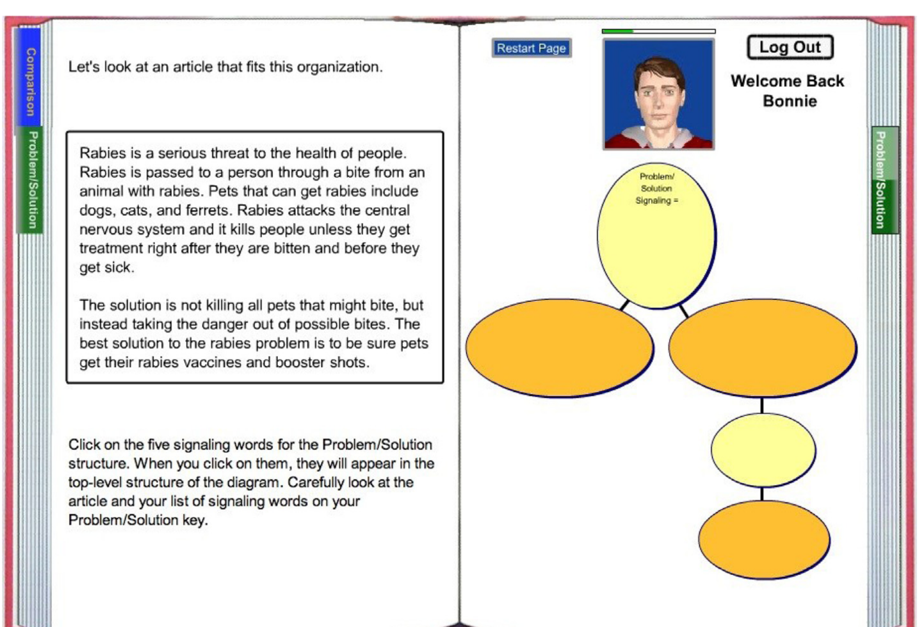

Fig. 2. ITSS interface with instruction about combined top-level structures of problem-solution with comparison.

Table 1
Categories of disengaged responses of relevant to off-task responses in ITSS.

Categories of disengaged behaviors	Definition	Antonyms/desired student behaviors	Examples
Off-task behavior outside the IT system (e.g., Baker, 2007; Baker & Rossi, 2013)	Behaviors that occur outside of ITS and that are irrelevant to instruction in ITS.	Time on task	Behavior not involved with the system or learning task, including off-task conversation, solitary activity (e.g., surfing the web) or inactivity
System active off-task responses ("off-task") Off-task behavior interacting within the ITS in ways not geared to learning the goals of the instruction (e.g., Rowe et al., 2009), which is similar to Without Thinking Fastidiously (Wixon, 2013; Wixon et al., 2012)	Student responses in ITS have no relationship to intended learning task. Playing with learning systems instead of diligently working on system's learning tasks.	Diligence working in ITS	Explores areas of micro-world irrelevant to ITS tasks, such as going places (i.e., climbing buildings) or doing things not necessary or helpful to the goals of instruction. Changing variable values many times without running the on-task related simulation. **ITSS**: Nonsensical letters (kjjjllllxx); lines or numbers (||||; 777); comments to I.T. ("Hi it." "I'll kill you if you don't shut up and I'll kill your dog too."); off-task gibberish ("cats + dogs = gerbils"); typing I.T.'s instructions (Write ALL you remember about the article.); irrelevant statements ("i no a song that get onevery bodys nerves."); smart-alecky comments ("Olympic athletes [text topic compares 3] khangas khan, godzilla, and yankee doodle ate poop and peed on the staue of puberty.")
Gaming the system (e.g., Baker et al., 2004; Muldner et al., 2011)	Trying to be successful in ITS by exploiting attributes of the system instead of mastering the system's learning goals.	Diligence in gaining deep understanding by thinking through the material.	Rapid guessing to find correct answer. Asking for hints or clicking on them rapidly to find a solution. Strategic ways to get answers without learning strategies and knowledge central to ITS learning goals. **ITTS**: Little to no response or off-task responses through trials and ignoring feedback (hints) until main idea is modeled, and then paraphrasing the answer.

Because Intelligent Tutoring systems have been successful in enhancing learning (Koedinger & Aleven, 2007), many have overlooked some of the problems that are inherent to these systems, such as off-task behaviors in and apart from the system and gaming. The preponderance of research examines gaming. Baker, Corbett, Koedinger, and Wagner (2004) explained that some students use help, feedback, and support via scaffolding from the software to avoid learning rather than aids to promote learning. Baker et al. define "gaming the system" as consisting of "behavior aimed at obtaining correct answers and advancing within the tutoring curriculum by taking advantage of regularities in the software's responses – systematically misusing the software's feedback or help instead of actively thinking about the material." Muldner, Burleson, Van de Sande, and VanLehn (2011) considered gaming for their gaming detector to include a) ignoring hints, b) copying hints, c) guessing when the tutor points out an incorrect response and the student quickly generated another incorrect response, and d) lack of planning exhibited seeking and using hints.

Due to characteristics of the reading comprehension tutor, ITSS, there are only a subset of the "gaming" responses investigated in past studies that could be relevant to the tutor–student interactions in ITSS. First, the primary tutor–student interactions in ITSS will be described. Next, the types of "gaming" identified by two prominent research teams publishing about this construct will be delineated and contrasted to the subset of off-task response in ITSS logs that might be identified as "explicit gaming trials," but appear to better match the off-task responses. Rowe et al. (2009) narrowed the focus of off-task behaviors in their study to match the disengagement problems within their narrative-centered learning environment, rather than looking at the more frequently studied gaming behaviors. They were particularly interested in student behaviors related to locations or objects not directly related to solving science mysteries in a created narrative world. We also narrowed the focus of the off-task behaviors to study explicit off-task responses submitted to the on-line tutor rather than expected on-task responses related to reading comprehension instruction.

1.1. Description of ITSS and off-task behaviors within ITSS

The first 12 lessons of ITSS describe the structure strategy and how to use it with comparison (compare & contrast) texts. After completion of the comparison lessons, ITSS presents 12 problem-and-solution lessons, which also include review of the comparison structure and integration with the problem-and-solution structure in one text (see Meyer et al., 2010 for a complete description of ITSS lessons). The next set of lessons in ITSS focuses on the cause-and-effect text structure with 16 lessons along with review and integration with the two prior structures taught in ITSS; only 45% percent of the students in the current study progressed through the cause-and-effect lessons. Fewer students worked through the next sets of lessons that covered the sequence structure followed by the description structure. The first three structures covered in ITSS are less familiar and more difficult than structure appearing later about sequence or descriptions (e.g., Ray & Meyer, 2011). However, when understood comparison, problem-and-solution, and cause-and-effect text structures provide more organization for encoding and retrieval from memory (Meyer & Freedle, 1984).

The ITSS lessons are self-paced. An intelligent tutor, an animated pedagogical agent called "I.T.," provided the feedback for all students. The instruction was written so that students would work in a particular ITSS lesson for the number of trials needed to master the content. During the first 2½ months for the current study's participants, this policy was followed. Unfortunately, constraints of scheduling and objections of 7th grade students told their work was incorrect resulted in limiting the number of trials per task in a lesson to four trials per main idea task and three trials per recall task. So nearly half way through the lessons, some students probably figured out how many trials were left before moving on to another task in the lesson or the next lesson after the final recall task. Current adaptations of ITSS have gone back to unlimited trials for the tasks with more modeling by I.T. in visual and auditory forms.

In the elaborated feedback condition in the present study, hints were provided with individualized feedback based on on-line performance. For example, the typical scenario for students not exhibiting mastery of the structure strategy for the main idea/summary task in ITSS lessons involved four trials. For performance below criterion (60% of main ideas) on the first trial, students received feedback from I.T. similar to "You're doing well, but you need to add more information." The next unsuccessful trial yielded feedback from I.T. referring the student to a visual with I.T.'s modeled main idea statement; I.T. said, "Please read my main idea and correct your work." If the students' third trials also were unsuccessful, I.T. would say some variation of "Please read my main idea and add anything I have that you don't have. Work must be corrected properly or we cannot move ahead." Students' fourth unsuccessful trial referred them to get help from their teachers, an approach used by Roll, Aleven, McLaren, and Koedinger (2007) to reduce gaming. In the simple feedback condition, students only learned whether or not they were correct. Due to the vocal objections of seventh-graders, after the first 2½ months all students in both conditions after their final incorrect trial heard I.T. saying, "Let's move on to the next page," rather than "Your answer is incorrect."

Within the lessons I.T. models how to read and look for signaling words that can cue readers into finding overall text structures, such as "on the other hand" and "however" for the comparison structure. I.T. also models how to use text structure to produce a good main idea statement and recall protocol during the initial lesson(s) for each of the five text structures. The tutor–student interactions continue until the student has completed writing a main idea and recall for the passage. In the first half of the lessons on each text structure, the student types the main idea while the passage is visible on the screen and submits the response to I.T. to check. In the second half of the lessons on each text structure, the student types the main idea for the passage with the text removed so that memory is an issue. Note that contrary to the cognitive tutors in the "gaming" research domain, students do not request hints or help. Instead, hints are provided on the basis of deficiencies in the student's performance. Also, contrary to the cognitive tutors in the "gaming" research area, instruction focuses on a reading strategy to increase memory and understanding rather than problem solving in math or physics. A response that a student does not remember is not necessary an off-task response and may be memory failure without a strategic goal to avoid learning. In fact, typically about 3–5% of participants cannot remember anything on a pencil and paper recall tasks; for this study 3.5% of the participants turned in blank recall sheets for their first paper and pencil recall task.

Ideally, but rarely, the student's main idea response meets the criteria on the first trial. However, Jeff produced the following main idea statement (summary) on his first trial on the 11th of the 12 comparison lessons in ITSS. By then, he had received many prior lessons about how to write a good main idea statement.

> "This story is about the differences of potbellied pigs and chinchillas. Frost, chinchillas like to be picked up. they like to go to the bathroom in their cage. And are very hyper compared to potbellied pigs. the potbellied pig and the chinchilla are very different.
>
> Last, potbellied pigs hate being picked up. they like to go to the bathroom outside and are very lazy."

This fifth-grade student described by his teacher as high-average in reading and scoring on a standardized reading comprehension on the pretest as reading at the third-grade level, received a "Super Job" comment from I.T. (71% of main ideas possible) and moved directly from the first trial on the main idea task to the next task in the lesson (recall). Jeff progressed slowly and carefully through 17 of the ITSS lessons during the 6 months of instruction offered in three sessions for a total of 90 min a week. Jeff produced no gaming or any off-task responses. Interestingly, his post-test score on an alternative form of the standardized reading comprehension test showed him performing after instruction with ITSS at the 11th grade level. Jeff observed the modeling by I.T., worked diligently on the practices lessons, and utilized the affordances of the software to move from no mastery of the structure strategy for comparison texts on the pretest to mastery of the strategy for comparative texts on the post-test and 4-month delayed post-test.

Holly, a seventh-grader, also produced no off-task or gaming responses in ITSS, but she required the following three trials to meet the main idea statement criteria.

> Trial 1: "Potbellied pigs and Churlishness are very different. Potbellied pigs are kind of lazy, they are not hyper like the Churlishness. they go to the bathroom outside and the Churlishness go to the bathroom in the cage that they sleep in." (42% of possible main ideas with I.T. responding, "Good job using signaling, but please add more information.")
>
> Trial 2: "Potbellied pigs and Chillies are very different. Potbellied pigs are kind of lazy, they are not hyper like the Chillies. they go to the bathroom outside and the Churlishness go to the bathroom in the cage that they sleep in. The chilli's are very hyper and they follow their master around every room." (42% of possible main ideas with I.T. responding that Holly should examine I.T.'s modeled main idea.)

I.T.'s modeled main idea was "<u>Chinchillas and potbellied pigs were compared</u> on <u>being picked up</u> (Chinchillas = get excited to be picked up, Potbellied pigs = don't like to be picked up), <u>hyper</u> (Chinchillas = yes, Potbellied pigs = no) <u>bathroom</u> (Chinchillas = go in their cages, Potbellied pigs = go outside)". A main idea uses the text structure that can organize all the information in a text into a coherent whole; this can assist students in building a coherent memory structure rather than just remembering disorganized lists of information from text, a common approach for middle school readers (e.g., Meyer, Brandt, & Bluth, 1980). The main idea pattern for comparison text is ________________and ____________(two or more ideas) were compared on _______________, _______________, and _________________. Instantiated with the content from the Chinchillas versus potbellied pigs article, the simple main idea would be Chinchillas and potbellied

Table 2
Examples of main idea task responses counted as off-task responses versus not explicitly off-task for lesson 11 content displayed in Table 3.

Examples of off-task responses – fifth-grade student in top 3% for off-task responses enters a series of letters for answers until the number of trials reached for automatic move to the next task. No learning exhibited and student appears to lack learning goals.

Student	Time (s)	Trial number – off-task or not	Student's response
Emma	First try	Trial 1 – off-task	"kjjjjjfghhfddddddddddddgjhhhhhhhhhhhuteeurrbnvffffffffffffgh hhhhhhhhhhwwwwwwwwwwwwwwwwwllllllllllllll xxxxxxxxxxxxx" (Tutor (I.T.) responds, "Please read your comparison main idea pattern.")
Emma	78	Trial 2 – not off-task (could be student error in the erasing process; 1+ min)	Erased above response & submitted a blank response (I.T. responds, "Please read my main idea and correct your work.")
Emma	59 14	Trial 3 – off-task Trial 4 – Off-task – Student submits the same explicit response for a main idea in 14 s.	"Itirgkmjuivdffkvmudfvkdfjvmdfuvndifnvdjjnvjjhdfvnfvvnuhvdkjnvkdjnbyh Fgjbhfngbuhnfgjnbfjghnvbjhnfbjhnfgbjnfgjhnbvhjgbjhnvjvfjhgvbhgfjhbfgvbdjbnnbjfdhgdfjk hgmfbndgndrtkjgnhfdfjnbkcnbnmfhbjjfvnbjkvccnbncvjvbmvnbkcvnbcnbjcxn,fgbnjgfcbngf,nbkjcnb, ncc.vkjbngcjbgm,nvbjvnbmvnjbnfgjbngfjnbjgdfbfgnbgfnbkfb,gkfjgbnfcbgfjkvbfbghjvfbcnxfbvjfghk jbfgvbfbvffvkjhfbvhdbfkvjhbdfkjhvfnvkhjfnvfdjknvjbhfjkvnbfdkjhbnkjfnbjfbnfjdnvkjdfbnbknjdffv hjbxkcfjvbnfhjbvfbfjnbvjkbfjhdfkjnkjhsdkjhfdkjhvdjkgdjnjgvndfjkhvkjxcfhvkxfhgvkjfxdngv, jdfnkljgndfkjngkdjfghdfkfjgbfkjdhgdfjh" (I.T. responds in Trial 3 with "Please read my main idea and add anything I have that you don't have. Work must be corrected properly or we cannot move ahead.") (I.T. responds in Trial 4 with "Ask the teacher for help.") Trial 3 and Trial 4 do not look like "strategic gaming" because Emma does not paraphrase I.T.'s model presented after Trial 2 and Trial 3, but continues off-task responses.

Not off-task – fifth-grade student with no off-task responses in all main idea and recall tasks in the comparison lessons.

Student	Time (s)	Trial number & off-task response or not	Student's response
Aiden	First try	Trial 1 – not off-task (attempt at main idea)	"potbellied pigs and another pig and how they are different from each other" (28% correct; I.T. responds, "You're doing well with signaling, but you need to add more information."
Aiden	65	Trial 2 – not off-task	"potbellied pigs and another pig and how they are different from each other like the pot bellied pig is not active and the other one isn't" (I.T. responds, "Please read my main idea and correct your work.")
Aiden	7	Trial 3 – not off-task (student submits the same response rather than producing a gaming response by paraphrasing or copying I.T.'s model after Trial 2)	"potbellied pigs and another pig and how they are different from each other like the pot bellied pig is not active and the other one isn't" (I.T. responds, "Please read my main idea and add anything I have that you don't have. Work must be corrected properly or we cannot move ahead.") Due to 7 seconds on to resubmit Trial this could be classified as a "gaming" response, but it lacks the strategic effort to get the right answer fast because the answer in the form of the modeled main idea was presented in feedback after Trial 2 and then again after Trial 3.
Aiden	57	Trial 4 – not off-task response (after modeling by I.T. the student picks up Chinchillas and a detail, but is still confused about the two creatures)	"potbellied pigs and another pig and how they are different from each other like the pot bellied pig is not active and the other one isn't Chinchillas run around the house" (I.T. responds to Trial 4 with "Ask the teacher for help" because the student is still below threshold for success.)

pigs were compared on being picked up, hyperactivity, and bathroom habits. I.T. models for students in the elaborated feedback condition focus on a complete main idea with the options for each creature noted in parenthesis.

> Holly's Trial 3: "Potbellied pigs and Chinchilla are very different. Potbellied pigs are kind of lazy, they are not hyper like the Chinchilla. they go to the bathroom outside and the Chinchillas go to the bathroom in the cage that they sleep in. The Chinchillas are very hyper and they follow their master around every room. The potbellied pigs are very intelligent too." (71% of main ideas with I.T. responding good work.)

The recall is only completed after the student has carefully read the passage and then pressed the "finished reading" button. At that point, the passage is removed from the screen and the student has to write a recall based on his/her memory of the text. Students are advised to use their signaling words and reminded to check the organization of their recall. Students work at their own pace through the lessons.

Off-task data for this study were collected from the log files in the system for tasks of writing a main idea and a recall. Each trial attempted in these tasks was recorded by the ITSS system. Human scorers coded the data with 100% agreement because of the clear differences between off-task topics and main ideas or recalls from specific texts (e.g., explicit off-task responses: "itirgkmjuivdffkvm…." or "I like potatoe chips theyre. Yummerific" versus content related responses: "potbellied pigs are very intelligent pigs"). Thus, in contrast to scoring gaming that can be nuanced and challenging for human observers or computational detectors (Muldner et al., 2011), scoring these explicit off-task responses could be achieved with simple instruction as well as diligence examining coding the responses.

1.2. Comparison between the ITSS reading tutor and math/science tutors

A difference between tutoring software used with math or science materials (e.g., Baker et al., 2004; Muldner et al., 2011) and the ITSS reading comprehension tutor is that in ITSS students cannot ask for hints. Instead, hints come automatically to the student as part of elaborated feedback geared to the on-line performance of the student on the task. Another difference is that paraphrasing I.T.'s model response (copy and pasting is not possible in ITSS) is not seen as negative, but a way for less proficient readers to learn about using the structure strategy with more scaffolding. Meyer et al. (2010) found that jumps in competency (i.e., no parts of the problem-solution

Table 3
Example of scoring recall for lesson 11: criteria, feedback, and acceptable responses.[a]

Comparison (signaling)			
CHINCHILLAS[b] (main idea)			
	GET (main idea)		
		EXCITED[c] (main idea)	
		PICKED UP (main idea)	
		MASTERS (main idea)	
	HYPER (main Idea)		
	GO (main Idea)		
		BATHROOM (main idea)	
		CAGES (main idea)	
		RUN (detail)	
			FAMILYROOM (detail)
POTBELLIED PIGS (main idea)			
	NOT HYPER (main idea)		
	INTELLIGENT (main idea)		
	INDEPENDENT (main idea)		
	NOT LIKE (main idea)		
		PICKED UP (detail)	
	CHOOSE (detail)		
		AREA (main Idea)	
			OUTSIDE (main idea)
		BATHROOM (main idea)	
	CUTE (detail)		
	EAT (detail)		
		TREATS (detail)	
			LOTS (detail)
			PET (detail)

Criteria: 1 out of 1 comparison signaling words to get positive (60%) feedback on signaling; 10 out of 17 main ideas to get positive (60%) feedback on main ideas; 5 out of 9 details to get positive (60%) feedback on details.
Scoring with feedback example for signaling = 1 or 100% (+), main ideas = 7 or 41% (−); details = 2 or 22% (−): (a) Elaborated feedback: "You wrote a good comparison signaling word which is great, but your main idea was not quite right and you are missing some details. Check the pattern for the main idea to make sure you understand what is being asked for and rewrite your details including all that you can remember from the text." (b) Simple feedback: "Try again."

[a] Paraphrases and misspellings were programmed into ITSS and counted as correct for each of the ideas in the above scoring structure. Examples for "chinchillas" and "excited" are shown below.

[b] Credicted as correct for chinchillas: rabbits, chinchellsare, rodents, Chincillas, chechilla, chechila Chinnchillas, Chanchilas, chinchilles, chincilas, chinclas, chinchillies, chinchilees, chinchelles, chinnlas, chinclas.

[c] Paraphrases and misspellings credited as correct for excited: eager, feel good, like being, ecited, ectied, excitied, exicted, escied, excied, excitedly.

structure used on the pretest to competency on the post-test) occurred only for less proficient readers (reading below their grade level) if they were in the elaborated feedback condition with this extra scaffolding.

Potential similarities between ITSS in these math and science tutors are a) advancing within the tutoring curriculum by taking advantage of the software's responses, b) avoiding active thinking about the instructional material, c) ignoring hints, and d) quickly generating another incorrect response after receiving feedback about a prior incorrect response. An overall examination of the ITSS response logs suggested that few students who might be gaming the system in the sense of strategically getting a correct answer with minimal thought by exploiting aspects of the software (see Table 1 under gaming examples for potential gaming in ITSS).

For example, in ITSS students who entered off-task responses of nonsensical letters continued this disengaged behavior for all trials regardless of receiving a modeled correct main idea after the second or third incorrect trial (see Table 2 in Emma example). The same was true for students who submitted two consecutive blanks; that is, the blanks tended to be submitted on all four trials. Overall Aiden's responses in Table 2 are typical of the below-average reader in fifth grade. Due to the 7 s duration of trial 3, Aiden's quick resubmission could be classified as a "gaming" response, but it lacks the strategic effort to get the right answer fast because the answer was provided in the feedback after trial 2 and was ignored by Aiden. A pattern of results most like the patterns viewed in the gaming literature would be a quick response in trial 2 after feedback that the first trial response was incorrect. Next in this sequential gaming pattern for the elaborated feedback condition would be more time spent on trial 3 using the information in I.T.'s feedback, I.T.'s modeled response. If success was not gained on trial 3, then trial 4 would require more information paraphrased from I.T.'s repeated main idea and more time spent than in trial 2. Alternatively, quick responses could quickly be sent back to I.T. until the final chance for success was available in Trial 4 with longer times observed as the student paraphrased I.T.'s main idea. A repeated measures ANOVA examining type of feedback and time spent on trials was conducted with the repeated measure of seconds before submission of responses in trials 2, 3, and 4 main idea task. Data came from the 11th comparison lesson that compared pets (see Table 3). If the students had discovered constraints of the software system, it would surely be by the 11th lesson of the 12 comparison lessons. Findings showed no significant differences in time spent during trials 2, 3, and 4 ($p = .56$) and no significant interaction between time on trials and feedback condition ($p = .42$). Overall, this suggests that students were not quickly hurrying through the trials until the model main idea was displayed.

If "gaming" (i.e., hurry through the trials with minimal effort until I.T. models the main idea after the second or third trial and then copy the main idea for the third or fourth trial) occurs it should be in the elaborated feedback condition. Thus for the elaborated condition, the main ideas tasks in the last two practice lessons (after students were familiar with the number of trials and type of feedback provided for each trials) were minutely examined. We searched for a particular pattern of response times: quick (less than 45 s on trial 2; note trial 1 times were unavailable in the log) but slower on trial 3 or trial 4 (at least 1 min and twice the time as spent on trial 2) after I.T.'s modeled main idea was displayed. At the same time we looked for substantial increases in on-task content for trials 3 or 4, but particularly for responses in the participant's earlier trials (e.g., trials 1 and 2) that included off-task behaviors itemized in Table 1 but also blanks and "I don't know" statements. A random sample of fifth-grade students in the elaborated condition for the 9th and 11th lessons was examined; 4 out of 48 (8%) main idea patterns could be classified as following a strategic gaming pattern with lines, blanks, or a few words in trial 1 and a quick repetition in trial 2, and at least 10 times the duration in trial 3 or 4. One person exhibited the prototypical gaming pattern from the literature for the last practice lesson (lesson #11) with a blank on trial 1 and trial 2 (11 s to send the same blank response), 319 s spent after the modeled feedback following trial 2, and 110 s on the final trial adding another sentence but not improving the score of 50% of the main ideas. A search of the complete set of data logs for blanks did not yield any other protocols with blanks that followed this strategic gaming pattern. Also, students with responses claiming that they didn't know or didn't remember were checked; they did not follow the gaming pattern.

Muldner et al. (2011) pointed to the wide range in the literature of the prevalence of gaming from 10% of students to all students gaming the system with 40% showing more than average gaming. They also noted the variations in gaming prevalence with different populations and contexts. In our sample with a particular type of ITS in the domain of reading comprehension, we appear to have approximately 8% of the students attempting to use the set-up of the system to maneuver through the system somewhat successfully with the least effort possible and little learning. However, we would much rather have the students gaming the system in this way involving reading of I.T.'s modeled main idea to produce their own main idea than other options, such as repeated trials of lines, nonsense letters, smart-alecky comments, etc. These explicitly marked off-task trials, rather than gaming, are the focus of our study. In our sample 57.50% of the students were involved in submitting explicitly off-task responses within ITSS as exemplified in Table 1 (see row two column four).

ITSS Students submitting off-task responses (or blanks) might be considered gaming in the broadest sense if they realized that they could survive through the task or lesson and make it to the next one by submitting anything. They would not received "good" feedback from I.T. with accompanying visual showers of coins or confetti, but they would move through the lessons and perhaps appear to be working in the eyes of their peers, particularly the students who were typing something rather than just submitting blanks. For example, tabs on the ITSS interface (see blue or green tabs in the web version in Figs. 1 and 2) could show peers that they were making progress in the lessons. The preponderance off-task behaviors recorded in the ITSS data logs are best characterized by the off-task definition rather than the definition of "gaming" in the extant, more extensive literature on gaming.

Wixon (2013) hypothesized that work avoidance goals would be related to system active off-task behaviors; he reported a low positive correlation between the trait and Without Thinking Fastidiously (.14, but not statistically significant). Wixon also expected a negative relationship between his off-task measures and mastery goals and no relationship with performance goals. Baker et al. (2004) posited that gaming the system may indicate that a student approaches instruction with the software in terms of performance goals rather than learning goals. Goal orientations are often studied in the context of academic achievement (e.g., Ames, 1992; Ames & Archer, 1988; Dweck, 1986; Elliot & Church, 1997; Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002; Meece, Blumenfeld, & Hoyle, 1988; Pintrich, 2000; Wentzel, 1999). Learning goals focus on developing competence or mastery. In contrast, performance goals focus on doing better than others (performance-approach goals) or at least not looking worse than others (performance-avoidant goals; e.g., Elliot & Church, 1997). Wixon did not find significant correlations with any of the measures related to goals and his off-task measures. Overall, similar to Baker's (2007) work with gaming, Wixon did not find individual difference traits to correlate with his measure of off-task behaviors within ITS. However, we predict less explicit off-task responses for students with mastery goals and more off-task responses for those with performing avoidance goals and work avoidance goals.

Explanations for disengaged behaviors of students in intelligent tutoring systems have been categorized as student features or lesson features (Muldner, Burleson, van de Sande, & VanLehn, 2010). Student features have been classified in terms of affect (particularly boredom; e.g., Baker, D'Mello, Rodrigo, & Graesser, 2010; Rodrigo et al., 2008), knowledge deficits, emotional self-regulation (Sabourin, Rowe, Mott, & Lester, 2013), and student goal orientation (Baker, Roll, Corbett, & Koedinger, 2005). Muldner et al. (2010, 2011) reported that student features are better predictors of gaming than instructional features, but others have more support for instructional features (Baker et al., 2008). Baker, Corbett, and Koedinger (2009) pointed to the features of confusing hints and poor interface design. Muldner et al. (2011) explained that discrepancies related to attributing disengaged behaviors (i.e., gaming) within ITS to primarily student versus lesson features could result from different populations, contexts, domains of learning, and units of analyses. They call for further work in this area.

1.3. Context and predictions

In this study we provide further work on disengaged behaviors with 5th and 7th graders in the domain of reading comprehension. In contrast to most extant work with high school or college students learning to solve problems in scientific domains, we look at younger learners attempting to solve the problem of how a text is organized and how to strategically use the text structure to write a main idea statement and better recall concepts from texts. We tested whether 5th- and 7th-grade students could learn the structure strategy via ITSS in six months of training for 90 min a week spread over two or three days. We examined two feedback conditions (simple feedback: good or try again versus elaborated feedback: specifying what was correct in a student's response and what needed improvement with steps for revision). Students who received ITSS with elaborated feedback made greater gains on a standardized reading comprehension test than students who received ITSS with simple feedback (Meyer et al., 2010). In the current study we examine off-task responses for the Meyer et al. (2010) sample plus off-task responses from another elementary school and middle school with students who used ITSS under the same feedback design conditions.

In this study we identify which types of students tend to respond with off-task responses versus which students do not. Then based on a rich set of individual differences measures as well as on-line performance measures, we provide reasons why students produce these off-task responses in the system versus more diligently engaging in the learning activities afforded by ITSS.

In this exploratory study of prevalence and possible causes of explicit off-task responses, we expect to replicate prior studies pointing to the negative relationship between off-task behaviors/gaming and measures of learning both within the ITSS lessons and on post-tests (e.g., Baker et al., 2010; Muldner et al., 2011; Rowe et al., 2009). Individual differences in motivation, efficacy, and strategy use impact learning (e.g., Schunk, 1991) and were collected in this study about off-task responses. We consider three main types of learner variables: demographic variables (e.g., gender, SES, school, grade); strategy and skill variables related to the ITSS instruction (e.g., reading comprehension test scores, completeness of responses to main idea and recall tasks within ITSS, pretest measures of structure strategy use in writing main ideas and recalls); and motivation, attitude, self-efficacy, and goal orientation variables (e.g., work avoidance, learning/mastery and performance goals, motivation for reading, self-efficacy, interest in using computers for learning).

O'Donnell, Reeve, and Smith (2007) suggested that learning goals are more likely to develop from instruction matched to a student's skill level or enabled through scaffolding rather than instruction that is too hard or easy. Similarly, we hypothesize that off-task responses will be more prevalent for students who find lessons too difficult for their levels of skills or structure strategy competence. Also, due to boredom we expect more off-task responses for students with proficient initial structure strategy use with comparison text (high top-level structure scores on the pretest) while working on the first 12 lessons in ITSS about comparative texts.

Scaffolding is central to Vygotsky's (1978) theory and can be a critical asset of web-based tutoring by offering tips, pointers, hints, and assistance; these are attributes of the advanced feedback condition in ITSS, but not the simple feedback condition. Thus, we expect the design feature of feedback to reduced off-task responses. However, like the findings of Muldner et al. (2010) we predict that most disengaged behaviors in children using ITSS will result from individual differences in skills, motivation, and attitudes rather than our design variation. We expect for our learning context that off-task responses will relate to the students' skills and proficiency using the structure strategy in the lessons followed by their performance on measures of goals, work avoidance, and reading motivation. Also, in parallel with Rowe et al.'s (2009) findings we expect off-task responses to relate to test performance and gender effects with boys submitting more explicit off-task responses than girls.

2. Method

First, the sample is described. Next, the scoring system is described in order to clarify the learning tasks and feedback variable. Then, the materials, measures, and manner of identifying off-task responses are specified.

2.1. Participants

For this study 186 students (95 boys and 91 girls) from two elementary schools (105 5th graders) and two middle schools (81 7th graders) in suburban and rural areas surrounding two small northeastern U.S. cities were randomly assigned to the two types of feedback (simple or elaborated). Most students participated in ITSS during the regular school day, but others worked on ITSS during an after-school learning program. Ethnicity of the sample was 85% White, 10% Black, 2% Hispanic White, 2% Asian, and 1% other. As a measure of SES free or reduced lunch at that schools was calculated at 9% for one school district and 61% for the other. Particular site of the school did not impact number of off-task responses in any of the analyses.

2.2. ITSS scoring of responses and feedback

Iterations of ITSS software have used a Flash enabled web-browser, NET framework, an animated pedagogical tutor with a human voice (who models the use of the strategy and gives immediate feedback), an interactive practice system with flash activities, a multi-dimensional assessment scheme with tools like latent semantic indexing (LSI) and parsers, spell checking, synonym checks, Penn Treebank (Penn Treebank, 2005), and Wordnet (Miller, 2005). Figs. 1 and 2 show sample screen shots from the software.

Students using ITSS receive immediate feedback from the system based on their responses. The student responses are parsed and checked against the scoring tree stored as a linked list in the ITSS database. If there are no matches against the scoring tree additional processing is performed to check customize spelling and synonyms lists. The system scores for the categories of structure, signaling words, main ideas, and details. The ITSS feedback system selects appropriate responses from the database based on the question/task, trial, and thresholds for combinations of scores for signaling words, main ideas and details.

Table 4
Example of scoring for a student's performance trial on main idea task (no off-task responses).

Student	Lesson #11a, question	Try	Student's response	% Main ideas	% Detail	I.T. verbal feedback
Bob	Main idea	1	chinchellsare hyper and potbelles are smart. on the other hand, potbelles don't like being picked up and chinchelles do.in contrast chinchelles like going to the bathroom in there cage potbelles like going outside.	16	23	**Simple**: "Try Again." **Elaborated:** "You got the structure and signaling correct, which is great, but your main idea was not quite right. Check the pattern for the main idea and make sure you understand what is being asked."
Bob	Main idea	2	chinchellsare hyper and potbelles are smart. on the other hand, potbelles don't like being picked up and chinchelles do.in contrast chinchelles like going to the bathroom in there cage potbelles like going outside.	16	23	**Simple**: "Try Again." **Elaborated:** "You got the structure and signaling correct, which is great, but your main idea was not quite right. Check the pattern for the main idea and make sure you understand what is being asked." **& Help Box** for **Elaborated** feedback condition with I.T.'s modeled main idea. For an example screen shot see Fig. 7a.
Bob	Main idea	3	Chinchillas and potbellied pigs are very different. chinchills get excited to picked up my their masters. chinchillas are hyper compared to potbellied pigs chinchillas go to the bathroom in their cages and run around familyrooms with their masters. potbelles are smart and don't like being picked up. in contrast to chinchelles going to the bathroom in there cage potbelles like going outside.	60	47	Super Job!

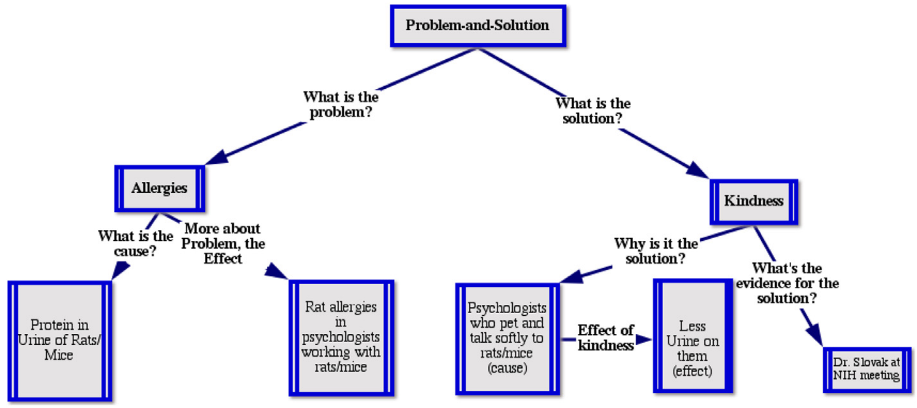

Fig. 3. Content structure of two-paragraph newspaper article.

The first step in this scoring approach is the top-down content structure analysis depicted in Figs. 3 and 4 (Meyer, 1975). Next, a detailed content structure is produced based on the rules of the semantic grammar of propositions (Grimes, 1975; Meyer, 1975). Data from hundreds of students were originally scored with an Excel program counting main ideas, details, and totals as well as hand entered signaling words, top-level structure scoring (highest levels in the structures shown in Figs. 3 and 4), and competence in using the text structure with the content of this text. Then, the scored data and original recalls were searched for acceptable misspellings and paraphrases of the capitalized ideas shown in Fig. 5. Misspellings and paraphrases are continually collected, updated with data from more users of ITSS, and programmed into ITSS for computer scoring. The next step was to remove the lower case labels of relationships among ideas except for the major rhetorical relationships organizing this article and shown in Fig. 3, the response: problem-and-solution structure and the related cause-and-effect structure. Signaling words for these relationships were scored by ITSS as well as the structure via signaling words (i.e., problem-and-solution). Thresholds, based on past data, are set for the percentage of correct ideas required for each of the four categories: structure, signaling words, main ideas, and details. Fig. 6 shows the ideas counted for these four categories. For example, criteria for a good main idea has been set in some of our work at 60% of the main ideas identified in Fig. 6. A human coder and supporting research team followed the scoring approach to provide relevant, individualized feedback or individualized sequences and types of lessons (Meyer et al., 2011). Reliability data between two human scorers and then between each human and ITSS scoring was compared for the main idea and detail scores. The percentage agreement between human scorers was 99% for main ideas and 97% for details, while agreement between a human and the ITSS scoring system for main ideas ranged from 95% to 97% and for details ranged from 71% to 78%.

Within ITSS students learn about specific main idea patterns useful to organize and remember the different text structures. For a detailed description of the ITSS instruction see Meyer et al. (2010).

Table 3 depicts an example of the scoring system used by ITSS in this study. A student's performance on the main idea task corresponding to Table 3 can be found in Table 4 along with scoring. Also presented are examples of auditory and visual feedback that would be provided in the simple and elaborated feedback conditions. In the elaborated feedback condition the Help Box provided by I.T. is similar to that shown in Fig. 7a. Fig. 7b depicts a student's off-task response submitted on a recall task and the recall Help Box provided by I.T. in our current ITSS interface. Fig. 8a and b shows the progressive scaffolding provided to this student followed through the main idea task and recall task (see Figs. 7 and 8). Meyer et al. (2011) further applied this scoring approach to individualize the sequence and difficulty of lessons to better match each student's on-line performance (see Meyer et al., 2011).

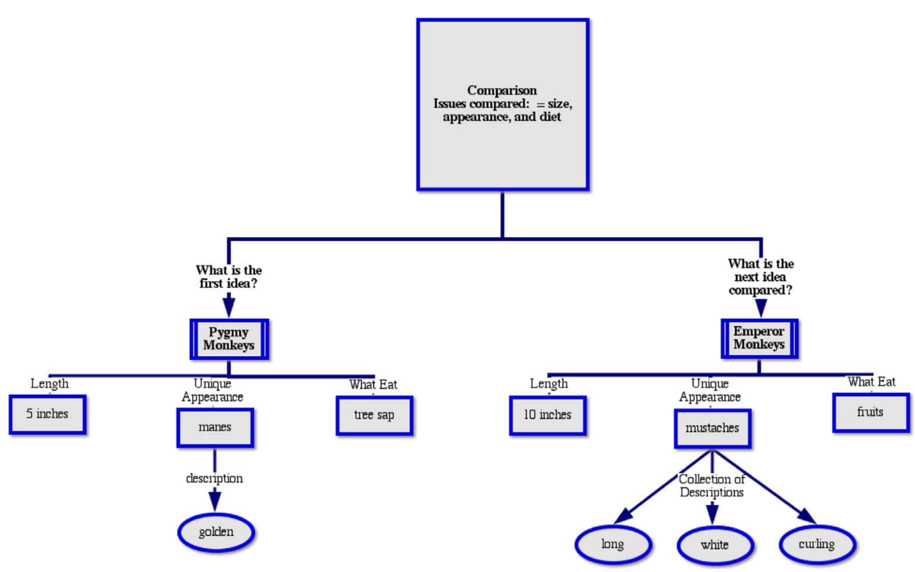

Fig. 4. Simplified version of comparative text examining differences between two primates on three issues with details shown in the ovals at the bottom.

```
1response
2 problem REAL HAZARD (can have problem and no solution)
3BECOME ALLERGIC TO CREATURES (get sick, diseases from rats; if 3 prob 2 also)
    4 manner
    5 OFTEN (usually)
    6 patient
    7PSYCHOLOGISTS (specific; must start with P & not look like another word)
          8 description: equivalent
          9 INVESTIGATORS (workers studying, playing, or learning w/
         10 description: attribution
         11 WORK WITH (do something with)
              12 patient
              13 collection
              14 RATS score if rats in isolation
              15 MICE
              16 range
              17 EXPERIMENTS (tests)
                     18 description: specific
                     19  RUNNING RATS IN EXPERIMENTS (studying in experiments with rats)
              20 setting: time (so much; lots; give 20, but not 21)
              21 HOURS A WEEK (some specific time, day & night)
    22 causation: explanation  CAUSES (is urine)
    23 ARE A REACTION
              24 patient, latter (for diseases)
              25 ALLERGIES (specifically allergy, not diseases)
              26 former
              27 URINE
                     28 description: specific
                     29 PROTEIN (IN URINE)
              30 range
              31 THESE SMALL ANIMALS (CREATURES/RATS/MICE (their, its)
32solution  recommended suggested
33 RECOMMENDED KINDESS TO (Goodness, good, gentle)
     34 benefactive
     35  collection
     36 RATS
     37 MICE
     38 instrument
     39 EXPERIMENTERS/PSYCHOLOGISTS/INVESTIGATORS (strict scientists types working with rats in experiments)
    40 agent (they say, but not passage says)
     41 Dr. SLOVAK Slavah
              42 description: specific
              43 ANDREW J. M. (not J.M. alone)
              44 description; attribution
              45 collection
              46 BRITISH
              47 PHYSICIAN doctor
     48 description: setting (where solution was reported)
     49 SPONSORED
               50 patient
               51 MEETING
          52 agent
          53 NATIONAL INSTITUTES OF HEALTH (Health Institute, Health place)
54 causation: explanation (for specific kindness & don't become allergic, then 54, 66)
55 causation: covariance: antecedent (need general kind or spec & 66
56 collection
57 PET (petting)
58 TALK TO
          59 manner
          60 SOFTLY
          61 agent
          62 PSYCHOLOGISTS (must try to write psychologists)
          63 patient
          64 RATS mice
65 causation: covariance: consequent
66 ARE LESS SPLATTERED WITH (don't become allergic, don't get diseases)
          67manner
          68 OFTEN (less likely splattered)
          69 range
          70 collection
          71 URINE
          72 PROTEIN THAT CAUSES ALLERGIC REACTION
```

Fig. 5. Content structure for top-level structure depicted in Fig. 3.

2.3. Materials

Measures used in the study included two forms of a standardized reading comprehension test counterbalanced over pretest and post-test, researcher designed tests (counterbalanced over time of testing), numerous questionnaires (pre- and post-test), cloze task on signaling words, and average performances on the main idea tasks and recall tasks in the ITSS lessons collected on-line during the lessons. We also received school records for the students' most recent scores on standardized reading tests administered by the school prior to our study. The lessons were self-paced on the computer and responses typed or clicked by the students. Questionnaires were administered on the computer, but testing of reading comprehension was collected with paper and pencil measures.

1 response (score if both solution and problem scored) structure
2problem HAZARD structure and signaling
3 ALLERGIC (Not allergy for paraphrase) main idea and signaling
5 OFTEN detail
7PSYCHOLOGISTS main idea
9 INVESTIGATORS main idea
11 WORK main idea
14 RATS (score if rats in isolation) main idea
15 MICE main idea
17 EXPERIMENTS main idea
19 RUNNING detail
21 HOURS detail
22causation(score if reaction or cause/effect signaling) main idea signaling structure
23 REACTION main idea signaling
25 ALLERGIES main idea
27 URINE main idea
29 PROTEIN main idea
31 ANIMALS detail
31.2 SMALL detail
32 solution recommended suggested structure, signaling, main idea
33 KINDESS main idea
36 RATS main idea
37 MICE main idea
39 EXPERIMENTERS main idea
41 SLOVAK detail
43 ANDREW detail
46 BRITISH detail
47 PHYSICIAN detail
49 SPONSORED detail
51 MEETING detail
53 HEALTH detail
53.2 INSTITUTES detail
53.2 NATIONAL detail
57 PET main idea
58 TALK main idea
60 SOFTLY detail
62 PSYCHOLOGISTS main idea
64 RATS main idea
66 LESS main idea
68 OFTEN detail
71 SPLATTERED main idea
72 URINE main idea

Fig. 6. Ideas labeled for four categories to be scored: structure, signaling, main ideas, and details.

2.3.1. Standardized reading comprehension test

The *Gray Silent Reading Test* (GSRT; Wiederholt & Blalock, 2000) was administered by the investigators and is a multiple choice reading comprehension test with a group administered option, reliable alternative forms, and testing of deep comprehension processes that include finding the main idea and reasoning with the text's main idea. The GSRT is designed to test readers 7 through 25 years of age all with the same test. Average alternate-form reliability was reported in the test manual at .85 (i.e., .87 for 10-year olds) and delayed alternate-form reliability was reported at .83. Coefficient alpha reported for forms A and B were .95 and .94, respectively. The school districts provided scores from the *Stanford Achievement Test* (10th Edition, 2003) Reading Comprehension subtest scores (SAT scores), administered in fourth and sixth grades.

2.3.2. Experimenter-designed reading comprehension tasks

Three equivalent forms of the tests were counterbalanced and administered before the students started ITSS, immediately after completing the program, and at a delayed follow-up. The problem-and-solution set of passages have 98 words, 72 idea units, and equivalent scores on traditional measures of readability and aspects of text coherence (see Meyer, 2003 and Fig. 3). They each present a problem and a solution on topics of rats, dogs, or cats; the article about rats was an authentic newspaper article (see Meyer & Poon, 2001). After reading the passage, students were asked to do a recall task, in which students placed the text in an envelope so they could not refer back to the text, and then wrote all they could recall from the passage. A similar set of three passages was prepared for the comparison structure (see simplification of structure in Fig. 4): pygmy versus emperor monkeys; leatherback versus hawksbill turtles; and emperor versus Adelie penguins. Each passage comparison has 128 words, 15 sentences, and 96 idea units. Three tasks were designed for these passages. First task focused on signaling with four blanks in the two-paragraph texts that could be filled with comparison signaling words (e.g., "Unlike"). The inter-rater reliability coefficient for the signaling test was .98. Reasons for only testing comparison signaling words on the signaling test included the 100% completion rate of lessons focusing on comparison signaling words and time constraints for testing within the schools. The second task was a main idea task where students were asked to write a main idea of no more than two sentences in length with the text available to examine. The final task was a recall task, which was the same task as for the problem-and-solution passage.

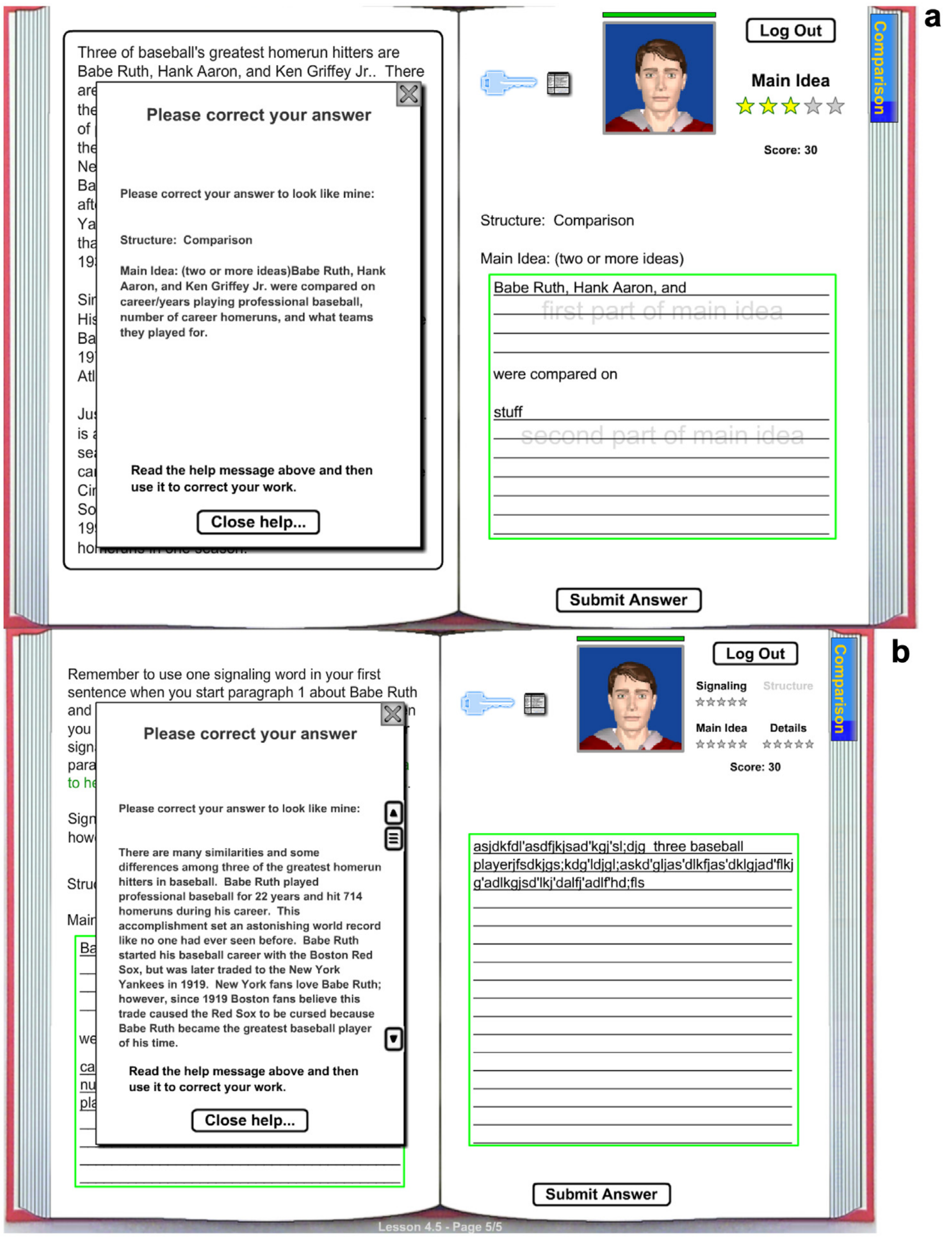

Fig. 7. Help boxes provided after several trials with poor student responses for main idea task (a) and recall task (b).

2.3.3. Goals orientation and work avoidance

Items to measure performance and mastery orientation were taken from *Patterns of Adaptive Learning Scales* (Midgley et al., 2000) with reported reliabilities from .74 to .89. Additional items were included to measure work avoidance (Meece et al., 1988) with reported reliabilities in excess of .8. The scales listed below were put on one questionnaire with a 5-point scale; items from a particular scale were interleaved rather than grouped together by type of scale on the questionnaire. For example the work avoidance items were presented as items 1, 6, 11, 13, 15, and 19. Internal consistency for students in our sample, measured by Cronbach's alpha statistics (1951), was .89.

To measure learning goal orientation, the Mastery Goal Orientation Revised Scale from the PALS was used. The scale has a reported coefficient alpha of .85 (Midgley et al., 2000). All five items from the scale were used. Only slight modifications were made to items to provide a context for the structure strategy work instead of school in general. For example, the original PALS item was "*One of my goals in this class is to learn as much as I can*." We modified the item to state, "*One of my goals when I do this work is to learn as much as I can*." Students were instructed to rate on a 5-point scale their agreement with the statements in terms of what they thought about working with ITSS. Scores on the measure range from 5 to 25, and Cronbach alpha reliability (Cronbach, 1951) for our sample was .91.

The Performance-Approach Goal Orientation Revised Scale from the PALS was used to measure learners' performance-approach tendencies. The Performance-Approach Goal Orientation Revised Scale consists of five items with a reported scale reliability of .89. All items were used and were only slightly modified to provide context for the structure strategy. An example item is "*One of my goals is to look smart in comparison to the other students*." Scores on the measure range from 5 to 25, and Cronbach alpha reliability was .87 for our sample.

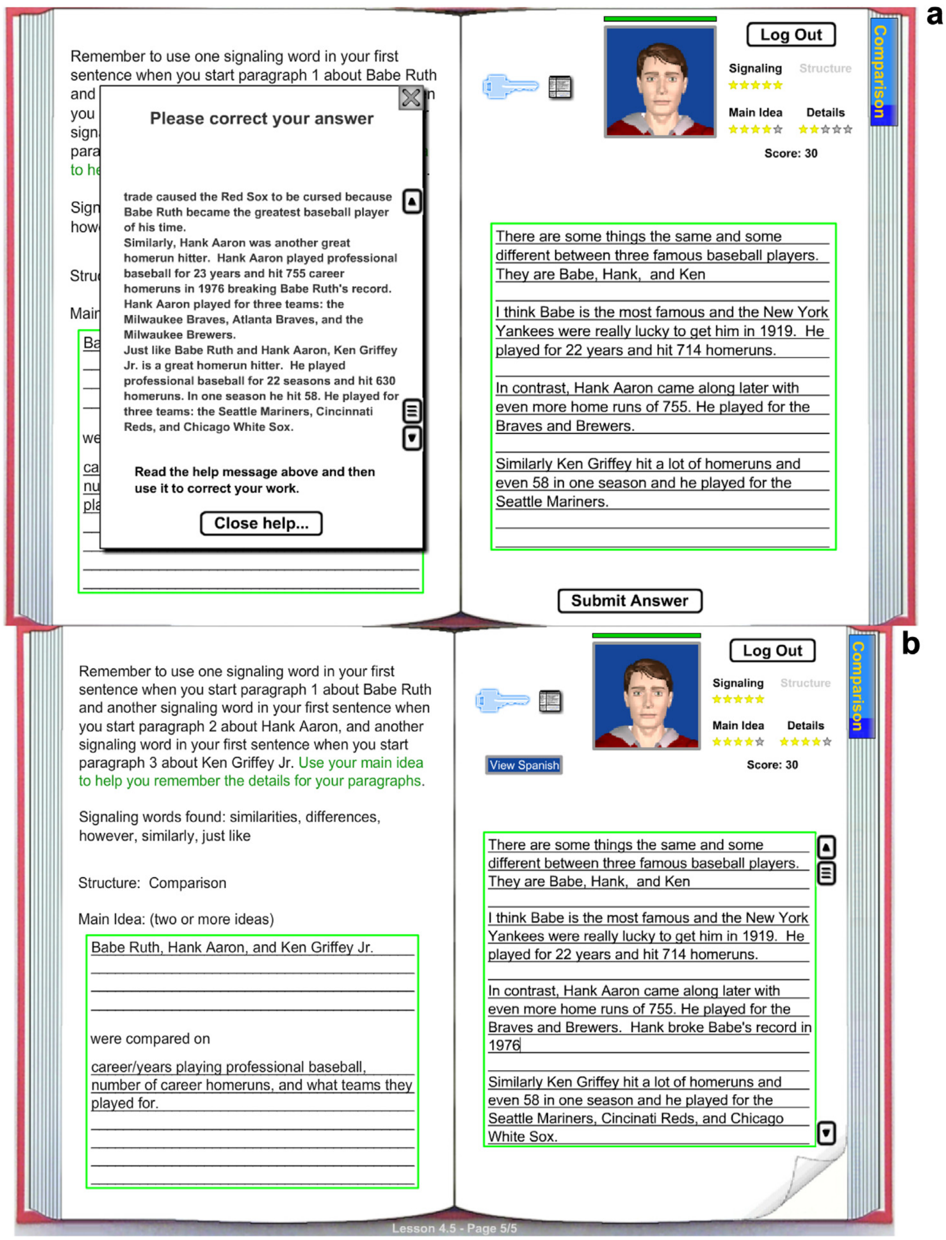

Fig. 8. I.T. says, "You are doing well, but need to add more details" (a) and after student meets 50% criteria for signaling, main ideas, and details "Wonderful Work!" (b).

We also included all of the items from the PALS Performance-Avoidance Goal Orientation Revised Scale. These four items have a reported scale reliability of .74. An example item is "*One of the goals is to avoid looking like I have trouble doing this work*." Scores on the measure range from 4 to 20, and Cronbach alpha was .87 for our sample.

Additionally there were six items on measure work avoidance scale (Meece et al., 1988). An item example is "I just want to do enough of this work to get by." Cronbach alpha was .88 for our sample.

2.3.4. Computer use and opinion questionnaire

A 21-item (5-point Likert scale) was used to query computer attitudes and preferences for reading from print versus computer (Krauss & Hoyer, 1984; Meyer & Poon, 1997; Meyer et al., 2002). An example item from this measure is "I like the idea of learning from the computer." Internal consistency for students in our sample, measured by Cronbach's alpha statistics (1951), was .85.

2.3.5. Self-efficacy scale

To measure self-efficacy, a 23-item (4-point Likert scale; reported reliability = .86) questionnaire was administered (Sherer et al., 1982). An example item is "When I decide to do something, I go right to work on it." Internal consistency for students in our sample, measured by Cronbach's alpha statistics (1951), was .86.

2.3.6. Motivation to read profile

Students used a 4-point Likert scale to rate 20 items on the *Motivation to Read* profile (Gambrell, Palmer, Codling, & Mazzoni, 1996). A sample item is "Reading is _____" with four options ranging from "very hard for me" to "very easy for me." Gambrell et al. (1996) reported reliability coefficients from .68 to .82; Cronbach's alpha statistics this study was .89.

2.3.7. Interest in ITSS texts

Students also completed a 24-item questionnaire where they rated their interest in reading topics about some of the texts used in ITSS on a 5-point. An example item is "I like to read articles about chinchillas and potbellied pigs."

2.4. Tasks for data about off-task response in ITSS and assessment of work in lessons

The first 12 lessons of ITSS describe the structure strategy and how to use it with comparison (compare & contrast) texts. All 186 participants worked through these lessons. Their off-task responses and work on other measures are the primary data for this study. After completion of the comparison lessons, ITSS presents 12 problem-and-solution lessons, which also include review of the comparison structure and integration with the problem-and-solution structure in one text (see Meyer et al., 2010 for a complete description of ITSS lessons). Only 148 students worked through these lessons. For the 148 students who completed both sets of lessons the correlation between the students' off-task responses from the comparison lessons and problem-and-solution lessons was .69 ($p < .0005$). The next set of lessons in ITSS focuses on the cause-and-effect text structure with 16 lessons along with review and integration with the two prior structures taught in ITSS. Only 84 students in the sample worked through the cause-and-effect lessons. Due to the reduced number of participants these data will not be examined in depth in this study, but instead used for cross-validation.

Within the lessons I.T. models how to read and look for signaling words and asks the student to practice doing the same. The interactions continue until the student has completed writing a main idea and recall for the passage. The student types the main idea while the passage is visible on the screen. The recall is only completed after the student has carefully read the passage and then pressed the "finished reading" button. At that point, the passage is removed from the screen and the student has to write a recall based on his/her memory of the text. Students are advised to use their signaling words and reminded to check the organization of their recall. Students work at their own pace through the lessons. Off-task response data for this study were collected for the tasks of writing a main idea and a recall. Each trial attempted in these tasks was recorded by the ITSS system (for an example see Table 3).

2.4.1. Assessment of work in lessons

The total number of lessons completed in ITSS probably depended on many factors, such as care in completing work in lessons, absence from class, computer access problems during the first few months, and so forth. Work in the lessons collected on-line was used as a rough estimate of whether students were attending to feedback and revising their work in the ITSS lessons. Using the scoring procedure shown in Table 3, the number of correct responses was calculated for each task completed by a student in ITSS. The running total of these performances can track a student's on-line progress. The average number of ideas scored as correct by ITSS per attempt/trial for each student within a lesson was examined in this study. These scores were added up for the 12 comparisons lessons and served as indicators of work in the ITSS lessons.

The first measure focused on quality and quantity of work for the main idea task. The average number of main ideas scored as correct by ITSS per trial for the main idea task was summed across completed comparison lessons.

The second measure about work in ITSS lessons examined the recall tasks. The average number of total ideas scored as correct by ITSS per trial for a recall task was summed across completed comparison lessons.

2.4.2. What counted as off-task responses submitted in ITSS

Tables 1 and 2 provide examples for responses counted versus not counted as an off-task response submitted within ITSS. Each try recorded for the main idea task and recall task was recorded by the ITSS system and then coded later by a human scorer as an off-task response or not. Off-task responses were coded as 1 while those listed as not explicitly off-task were coded as 0 for a trial. A random selection of 10% of the trials indicated rater agreement for categorization of the presence or absence of off-task responses was 100%. The number of off-task responses for the comparison lessons was totaled for each student as well as their problem-and-solution off-task totals and cause-and-effect totals, if the students reached these lessons. Lines, nonsensical letters, gibberish (see Fig. 7b and Tables 1 and 2), or nonsense words were counted as off-task responses for each trial. Also, irrelevant statements for the task, "smart-alecky" comments to IT (note "stuff" response in Fig. 7a and "staue of puberty" in Table 1), or copying I.T.'s instructions for the task were counted as off-task response. Blanks (no response) or stating "I don't remember," or similar paraphrases about not knowing were not counted in our measure of explicit off-task responses. In summary, our off-task measure focused on the number of explicit off-task trials produced by students working with an on-line reading comprehension tutor.

3. Findings and discussion

3.1. How much explicit off-task behavior?

The distributions in Figs. 9 and 10a and b show the extent of off-task responses submitted by students in ITSS. Most students were not repeatedly submitting off-task responses. Nearly half of the students never submitted explicit off-task responses to the ITSS tutor.

3.2. Did students submitting off-task responses learn less?

The off-task scores were dichotomized into students who submitted off-task responses and students who did not. The average performance scores on main idea and recall tasks within the comparison lessons that students with off-task responses completed was

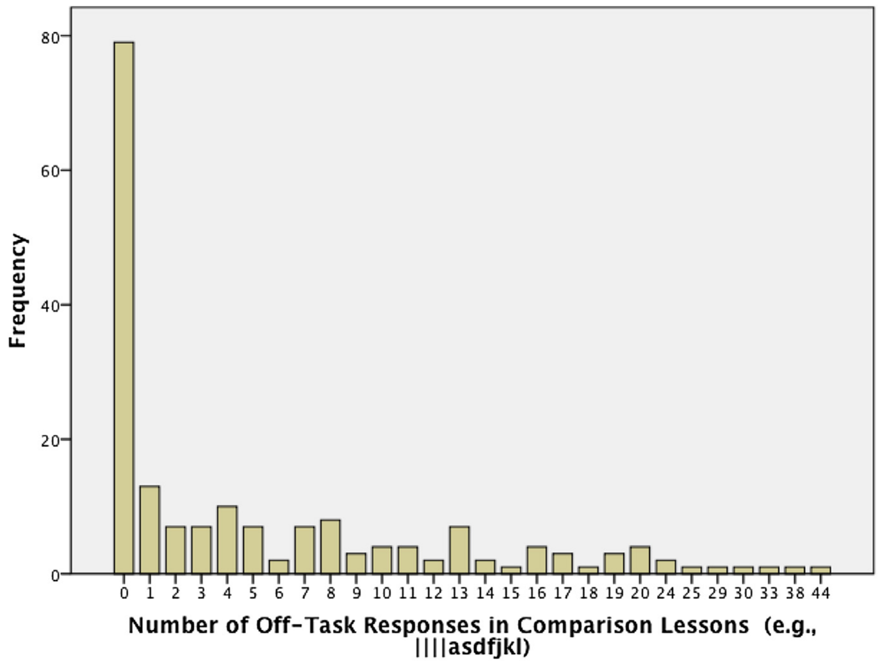

Fig. 9. Number of off-task responses (0 to 44) submitted by students ($N = 186$) within the ITSS comparison lessons.

significantly lower than the average performance level of students who did not submit off-task responses (Off-task: $M = 89.97$ [SD = 53.79], Not Off-task: $M = 153.13$ [SD = 40.43], $t(178.95) = 9.02$, $p < .0005$). Thus, the quality of work completed by students submitting off-task responses within ITSS lessons was lower than the work quality of students who did not submit such responses.

In order to test for learning from ITSS, post-test scores from the main comparison dependent variables (top-level structure and total recall scores and cloze signaling test) were examined for students who did or did not submit off-task responses after co-varying their pretest performance on these same scores. Results from a MANCOVA conducted on these data indicated that the students submitting off-task responses learned less in the ITSS lessons than students who did not (Wilks' Lambda = .89, $F(3, 141) = 5.78$, $p = .001$). Univariate analyses indicated the same statistically significant pattern of findings for each of the specified comparison variables (top-level structure: $F(1, 143) = 5.03$, $p = .026$; total recall: $F(1, 143) = 15.69$, $p < .0005$; cloze signaling test: $F(1, 143) = 9.09$, $p = .003$). These findings, which show less learning for students submitting off-task responses to I.T. in ITSS, replicate past work both on gaming (Baker et al., 2010; Muldner et al., 2011) and off-task responses (Rowe et al., 2009).

3.3. Who games and why?

Producing off-task responses leads to less learning. This study looks at which students provide such responses in ITSS and potential reasons for these off-task responses in order to decrease future off-task responses and increase on-task behavior and learning.

In order to answer these questions, Classification and Regression Trees (C&RT) and logistic regression analyses were conducted to give different perspectives and a richer view of the data since C&RT can point to nonlinear relationships plus provide an exploratory and visual display of the data. These analyses were aimed primarily at examining learner characteristics that relate to off-task responses within ITSS.

3.3.1. Decision-tree analysis for ITSS comparison lessons

C&RT modeling within AnswerTree 3.1 (SPSS, 2001) was chosen to construct the decision tree that shows clustering of similar participants. This method is considered an exploratory data analysis technique that examines relationships between a dependent qualitative or quantitative measure and a large number of potential predictors. C&RT modeling was chosen because it is viewed as being analogous to forward stepwise regression analysis. Also, it is often useful in summarizing data. C&RT has several advantages in examining complicated datasets. A key advantage is that the technique can be used with dependent and independent variables that are nominal, ordinal, or interval. Also, partial data can be utilized because missing values can be estimated using other variables. Additionally, the dependent variable does not need to fit a normal distribution (see in Fig. 9). C&RT can handle this lack of normality for the off-task responses variable while normality is an assumption for regression. Splitting the variable into a dichotomy or category would have lost considerable information because of the range of the off-task variable. Thus this richness in the data was retained with the C&RT analyses.

The main C&RT analysis began with an initial trunk containing all participants for 5th and 7th grade; these participants completed all of the ITSS lessons about the comparison text structure. The predictors were assessed via the C&RT analysis approach; a decision-tree analysis is split based on the predictors that best discriminate on the dependent variable with the first split being the variable chosen as the "most significant predictor." The resultant tree diagram displays a series of groups that are maximally different from one another on the dependent factor. An important characteristic of C&RT is that each optimal split is binary and leads to two new branches in the tree with each child node purer in respect to the dependent variable (extent of off-task responses) than the super-ordinate parent node. Thus, all cases would have the same value for off-task responses for a completely pure node. As a simple example, gender and work avoidance were entered as predictors of off-task responses in the problem-and-solution lessons. To keep the example tree simple the tree depth was set to 2 rather than the default

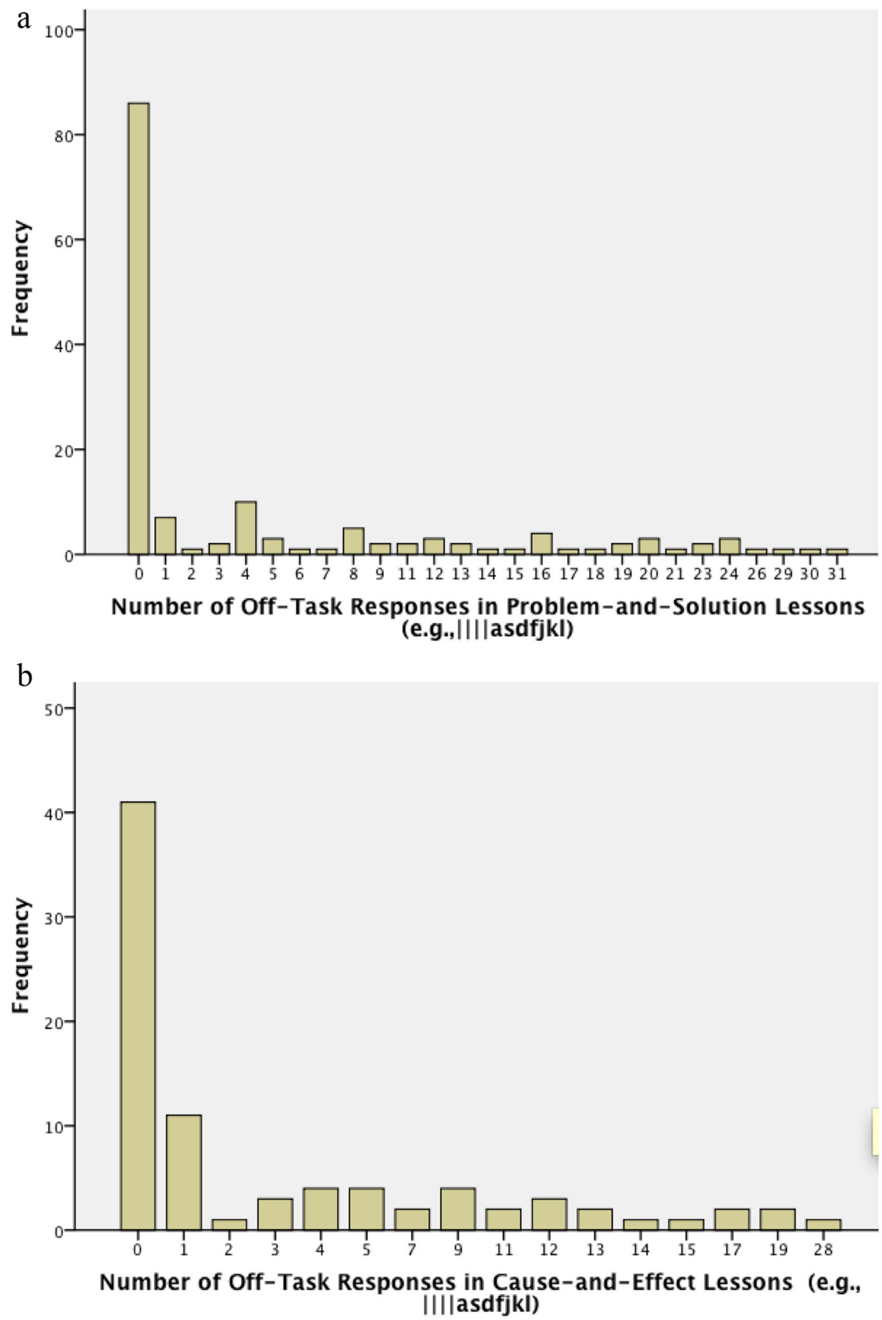

Fig. 10. (a) Students ($N = 148$) submitting off-task responses in the ITSS system within the problem-and-solution lessons. (b) Students ($N = 84$) submitting off-task responses within the ITSS system in the cause-and-effect lessons.

setting of 5 in SPSS Answer/Decision Tree. The first split, the most significant predictor, was gender. Females ($n = 77$) were placed at a terminal node with an average of 1.3 off-task responses, which could not be split further to improve the performance of the tree. However things worked differently for males. There were 8.5 off-task responses on average submitted by the 71 males; a split on work avoidance for the males led to two new branches of greater purity in respect to off-task behaviors than that of the super-ordinate (parent node) males. The binary split between males with work avoidance scores greater than 25.5 was split from males with scores equal to or smaller than 25.5. The 13 males with work avoidance scores greater than 25.5 on the new right branch had an average of 12.1 off-task responses, while the 58 males with work avoidance scores equal to or less than 25.5 on the new left branch had an average of 7.7 off-task responses.

For the C&RT analysis in this study, the maximum tree depth was set to 5, the default C&RT setting. Based on prior work with C&RT with similar sample sizes and number of predictors the parent nodes was set at 8 and the child node was set at 4; further divisions would probably be too minute to reasonably discuss or fit in one manageable figure. Additionally, the IBM SPSS 17 Decision Tree (SPSS, 2007) adaptation of Answer Tree allows the option of forcing the first variable into the prediction. This option was selected to test the feedback variable (simple versus elaborated) in C&RT analysis. Overall type of feedback was not a significant predictor of off-task responses in ITSS ((χ^2 1, $N = 186$) = .09, $p = .80$), and patterns produced after forcing the split were similar within each feedback condition. Thus, feedback condition minimally impacted off-task responses.

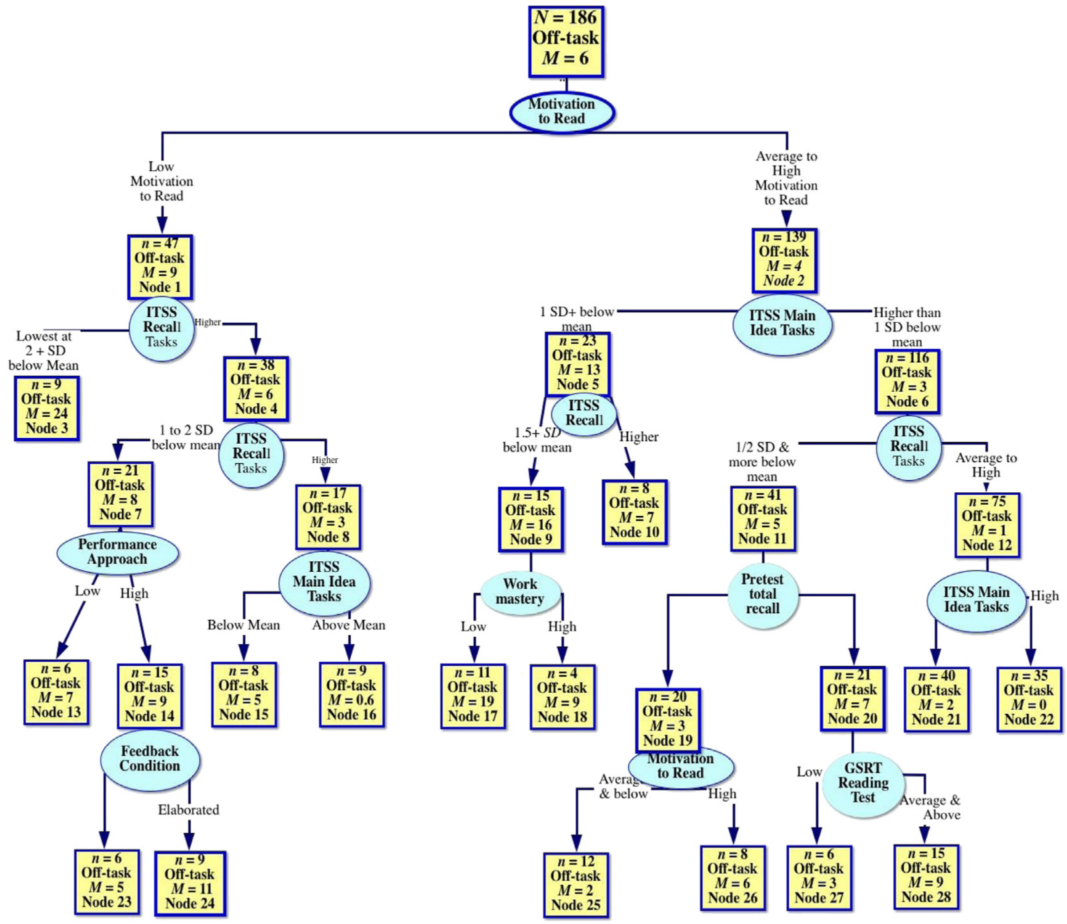

Fig. 11. Decision-tree analysis with all 19 variables entered including performance on main idea and recall task for the comparison lessons in ITSS ($N = 186$).

Variables that we predicted to be related to explicit off-task responses were selected as potential predictors in the decision-tree analysis. They included the seven reading strategy and skill variables: reading comprehension test scores on the GSRT as well as the Stanford Achievement Test, pretest signaling performance, pretest measures of structure strategy use and recall proficiency, and completeness of responses to main idea and recall tasks within ITSS. Additionally, they included seven motivation, attitude, self-efficacy, and goal orientation variables: work avoidance, mastery and performance (approach and avoidance) goals, motivation for reading, self-efficacy, interest in using computers for learning. There were four demographic variables: gender, SES, school, grade. Then, there were two variables related to design characteristics: feedback and the total number of lessons completed by a student in ITSS. ITSS was designed to be self-paced.

3.3.1.1. C&RT analysis entering all predictors. Eight of these variables were selected by the C&RT Decision Tree program to be used in constructing the decision tree (see Fig. 11). Some were used several times at different cut points in the decision tree. The identified important

Table 5
Gain and average off-task responses for 15 terminal nodes predicting off-task behaviors in the comparison lessons.

Node	*n*	Gain %	*M* Off-task	Description (motivation to read = MR, main ideas task in ITSS = MIT), performance-approach goals = PA
3	9	4.8%	24.44	Low motivation to read (MR) + low recall in ITSS lessons
17	11	5.9%	19.18	High MR, low MIT, low recall ITSS, lower work mastery
24	9	4.8%	11.33	Low MR, better than lowest recall in ITSS, higher PA goals, elaborated feedback
28	15	8.1%	9.00	High MR, higher MIT, lower recall in ITSS, higher pretest recall, higher GSRT reading comprehension scores
18	4	2.2%	8.75	High MR, low MIT, low recall in ITSS, high work mastery goals
10	8	4.3%	6.88	High MR, low MIT, lower recall in ITSS
13	6	3.2%	6.50	Low MR, better than lowest recall in ITSS, low PA goals
26	8	4.3%	5.63	High MR, higher MIT, lower recall in ITSS, higher MR
23	6	3.2%	5.00	Low MR, better than lowest recall in ITSS, higher PA goals, simple feedback
15	8	4.3%	4.88	Low MR, better than 1 SD below *M* recall in ITSS, less than average MIT
27	6	3.2%	3.33	High MR, higher MIT, lower recall in ITSS, higher pretest recall, lower than average GSRT reading comprehension scores
25	12	12.5%	1.92	High MR, higher MIT, lower recall in ITSS, lower MR
21	40	21.5%	1.83	High MR, higher MIT, higher recall in ITSS, lower than high-average MIT
16	9	4.8%	.56	Low MR, better than 1 SD below *M* recall in ITSS, better than average MIT
22	35	18.8%	.06	High MR, higher MIT, higher recall in ITSS, high MIT

predictors were **motivation to read** (twice – a division between students below average in motivation for reading versus those average and above and another division between students a standard deviation above the mean in motivation to read versus those scoring below that point), students' average performance on the **comparison main idea task** in the lessons (thrice – a division between students about a standard deviation below average versus average and above; a division of students above versus below the mean, and a division between students slightly above the mean versus students below that point), average performance on the **comparison recall task** in the lessons (four times – dividing students scoring over 2 + SD below the mean from those above; dividing students into two groups at about 1 SD below the mean; dividing low and higher students again at a point 1.5 SD below the mean; and dividing the 116 students at node 6 into two groups at a point about a half of SD below the mean), **pretest comparison total recall** (dividing below-average students on the comparison recall pretest from average and above students), **performance-approach goals** (dividing students slightly below the mean from those average and above), **feedback condition, work mastery goals** (dividing students above the mean from those below it), and GSRT **reading comprehension test scores** (dividing students scoring below average from those scoring higher). Fig. 11 displays the descriptive information for the tree including the number students in each node with their average (*M*) number of explicit off-task responses in the comparison lessons. Table 5 presents the results of the gain statistics in percentages and the average off-task score for the students at each of the terminal nodes in the tree. The proportion of the variance in explicit off-task responses in ITSS explained by the model outlined in the decision tree in Fig. 11 is 66%.

The decision-tree analysis provided some support for the hypothesized relationship between the predictor variables and explicit off-task responses within ITSS. We predicted that the motivation and skill variables would be more powerful than the design variables and this was supported. The motivation to read scale was expected to relate too off-task responses, but we were surprised by its prominence as the primary split at the root of the tree, which divided 47 students low in motivation to read from 139 students higher in motivation to read. The group higher in motivation to read displayed half of the off-task submitted responses as the group of 47 students with little motivation to read. The opposite relationship is found between nodes 25 and 26 involving only 20 students, where eight students scoring on average a standard deviation above the mean on motivation to read gamed the system more than 12 students scoring not as high on motivation to read. Looking down the nodes in the decision tree from node 2 to node 26, we can see that these eight motivated readers are grouped with lower groupings for average recall in the lessons and total recall on the pretest for the comparison text. Possibly the mismatches between their motivation and their performance on the recall tasks were causing some frustration and possibly more off-task responses to cope with these feelings (Sabourin et al., 2013). Overall two splits for motivation to read suggests a curvilinear relationship between "off-task responses" and motivation to read with this split between nodes 25 and 26 in opposition to the primary split on motivation. Although low motivation to read goes along with more "off-task responses" for most students, among the average and above students in motivation to read we see this subgroup where the pattern is switched. They appear to be students that do well with the scaffolding provided for the main idea task, but not as well on organizing their recall using the structures strategy on the free recall tasks in ITSS or the paper and pencil pretest. They may be motivated readers with memory difficulties, trouble self-regulating, or organizing their ideas in a free recall task. These students also could be bored (Baker et al., 2010) with a free recall task, but believe they are good readers and that reading is a valuable activity.

Another possible mismatch between instruction and skills may be evident at nodes 27 and 28; students at node 28 have higher than average GSRT reading comprehension skills and average or above pretest recall scores, but more off-task responses than students with lower skills at node 27. This may exemplify off-task responses as a result of boredom in the comparison set of lessons prior to the more difficult lessons that add and integrate other text structures.

As predicted, performance on the main idea and recall tasks related negatively to off-task responses. At the nodes closer to the root of the tree these predictors showed that higher performance in the lessons was related to less responses that were off-task. For example for the more motivated readers at node 2, the subset of 116 students with better main idea performance rarely submitted off-task responses while the 23 students with low main idea performance produced on average about six times the off-task responses as the group of 116 students. This dramatic pattern also can be seen for comparison recall performance in ITSS, which is evident by comparing node 3–4 and node 7–8 for the less motivated readers. Additionally, as seen in Table 5 for nodes 21 and 22, students with high motivation to read and consistently good performance drafting main ideas and recalls in ITSS rarely submitted off-task responses in the system. These students exhibit very little, if any, explicit off-task responses. Note that this is particularly the case for the 35 students with very high scores on the main idea task (see node 22 at the bottom of Table 5 and at the far right in Fig. 11).

Our predictions were supported by the relationship between work mastery goals and less off-task responses (see nodes 17 and 18). Our hypothesized link between work avoidance scores and off-task responses was not evident with the 186 students in the comparison lessons.

In examining nodes 7, 13, 14, 23, and 24 there is an interesting pattern of relationships with performance-approach goals, the feedback condition, and off-task responses. There are nine students in the elaborated feedback condition (see node 24) who follow the path in the decision tree from the root node with less reading motivation, low recall scores in the comparison lesson (albeit better than some), goals to perform well in front of their peers, teachers and perhaps I.T., but relatively high explicit off-task responses compared to six similar students in the simple feedback condition (see node 23).

We looked a two ways to cross-validate the tree shown in Fig. 11. First we examined all of the same inputted predictor variables as examined with the 12 comparison lessons again with the problem-and-solution lessons as well as the same tree depth, parent, and child node constraints. The problem-and-solution lessons have 80% of the same participants as used with the set of 12 comparison lessons. Also, we examined the same predictor variables, tree depth, parent nodes and child nodes with the cause-and-effect set of lessons with only 45% of the same participants as those used in the analysis of off-task responses in the comparison lessons. The five levels of the decision tree in Fig. 11 probably overly customize the splits of branches in the tree to the data set. We were interested to see splits with the motivation to read variable for all three analysis with different texts, tasks, and subgroups of participants as well as the variance predicted for the three tasks with varying subsets of learners. For the analysis with the problem-and-solution off-task responses the primary motivation split was still at equal to or less than 56.5 on the motivation to read scale versus above 56.5 on the scale. The next splits were again average quality of work in ITSS for the main and idea and recall tasks followed by again by the low motivation group with high performance approach producing more off-task responses. The variance accounted for by the decision tree was 67% with data from the problem-and-solution lessons. In the analysis of the data from the cause-and-effect lessons, motivation to read was also split at a score of 56.5 with students

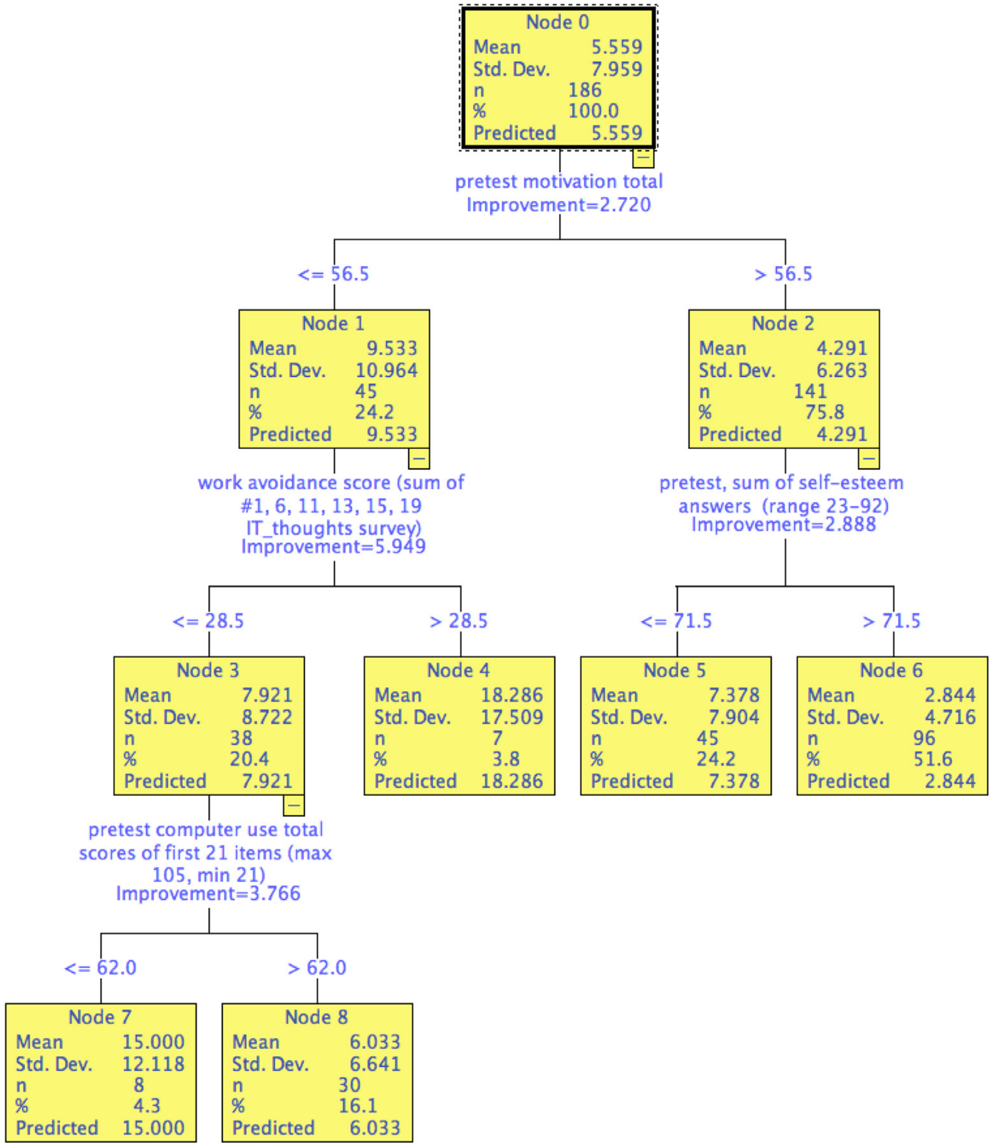

Fig. 12. Decision-tree analysis with 17 variables entered but without performance on main idea and recall task for the comparison lessons in ITSS ($N = 186$).

showing higher motivation to read producing less off-task responses. Average quality of work on the main ideas in ITSS also predicted off-task responses as did goals; 53% of the variance could be predicted from the decision tree.

Similarly we ran the cross-validation examination again but this time just putting in the variables in the model shown in Fig. 11. Again the tree was first split for motivation to read at the score of 56.5 and then by work in the ITSS lessons followed by work mastery/learning goals and at utmost branches feedback condition predicting off-task responses in the direction shown in Fig. 11. Off-task responses were predicted by lower motivation to read, poorer work in the ITSS lessons, and lower work mastery/learning goals. This very similar model as that shown in Fig. 11 predicted 46% of the problem-and-solution off-task responses. A similar result was found for the cause-and-effect lessons with the split for motivation again at score of 56.5 followed by a split for work in the lessons and at the lowest branches a reversed relationship for motivation to read. This pattern for motivation to read was similar to that found for the comparison lessons and much larger sample size. It points to the curvilinear relationship between motivation to read and off-task responses and why this variable may explain little variance in a linear data analysis.

The upshot of the decision-tree analysis is that most off-task responses come from low motivation for reading and poor performance on writing main ideas and recalls in the ITSS lessons. The question is, does the low performance in the ITSS lessons come from an inability to do the work in the lessons or a lack of will to work in the lessons? Basically, do off-task responses cause low performance in the lessons or does poor performance cause off-task responses? Motivation to read as the first separator of students in the decision tree in Fig. 11 suggests that the “will” to do the work may be a critical factor.

3.3.1.2. C&RT analysis excluding performance in lessons. To further investigate these questions, a decision-tree analysis was run again with the same predictors except that performance in the lessons was excluded from the analysis (i.e., comparison main idea task and comparison recall task). This analysis is shown in Fig. 12 and replicates the first split in Fig. 11 on motivation to read. Students with less than average motivation to read provided over twice as many off-task responses on average in the comparison lessons than students with higher levels of motivation. Next as seen at nodes 3 and 4 in Fig. 12, students nearly 2 SD above the mean (17.57 [SD = 6.17]) on the work avoidance scale responded on average with more than twice as many off-task responses ($M = 18.29$) than did students with low work avoidance scores ($M = 7.92$). For the latter group with lower work avoidance, those with more positive attitudes toward computers gave fewer off-task responses ($M = 6.03$) than those with more negative attitudes ($M = 15$). For students with higher reading motivation the only other separation was between students above and below average in self-efficacy. Students with higher self-efficacy scores submitted fewer off-task responses than those with lower self-efficacy/self-esteem (see nodes 5 and 6 in Table 6 and Fig. 12). The model in Fig. 12 shows fewer off-task responses through the influence of motivation for reading, self-efficacy about one's ability to work hard and overcome difficulties (Sherer et al., 1982), low thoughts oriented toward work avoidance, and positive attitudes toward computers. This model is clearer to understand through its simplicity but explains less of the variance in off-task responses than the model shown in Fig. 11, which included the two performance measures about main ideas and recalls in the comparison ITSS lessons. The motivation and attitudes model shown in Fig. 12 explained 20% of the variance in submission of off-task responses.

The biggest news from this C&RT model is that 96 of the students who rarely, if ever, submitted off-task responses to the system can be characterized by two motivation variables (see Fig. 12, node 6). They were motivated to read and high on a self-efficacy measure characterized by agreement to statements, such as "When I decide to do something, I go right to work on it." Prior to working on ITSS they were motivated to read and motivated to work on ITSS as a tool to increase their reading comprehension, and that is exactly what they did in ITSS without distracting themselves by submitting off-task responses or submitting off-task responses as an easy way out of work in the lessons.

3.3.2. Logistic regression analyses for off-task responses submitted in the comparison lessons

Logistic regression was used to examine the 19 measures initially examined with decision-tree analysis. The same seven measures of reading strategy skills used in the above decision-tree analyses were entered as predictors of explicit off-task responses versus no off-task responses (scores of 0) entered in a logistic regression analysis. Average of scores on the comparison main ideas tasks was the only significant predictor of off-task responses (Wald = 41.12, df = 1, $p < .0005$). The statistical model with scores from the main idea tasks average over all attempts correctly predicted 72.4% of no explicit off-task responses and 78% of some submitted off-task responses; Nagelkerke's $R^2 = .43$ (Nagelkerke, 1991).

Both the decision tree and the logistic regression analyses pointed to the importance of consistently good performance for the trials attempted on the main idea tasks across the first set of 12 lessons about the comparison text structure. Individualizing instructions, materials, and lessons sequence based on this variable may be able to prevent some of off-task responses. This variable is relatively simple to access and track on-line to modify the intervention in order to better meet the needs of students with a tendency to game the system.

A second logistic regression looked at the seven motivation, attitude, self-efficacy, and goal orientation variables with respect to their prediction of off-task responses. The only significant predictor of off-task responses was self-efficacy (Wald = 4.93, df = 1, $p = .026$). The statistical model with self-efficacy as the sole predictor correctly predicted 27.3% of no off-task responses and 86.2% of some submitted off-task responses [Nagelkerke's (1991) $R^2 = .06$]. The discrepancy between the decision-tree analysis and the logistic regression in regard to motivation to read relates to the curvilinear relationship between off-task responses and motivation to read. This curvilinear relationship was suggested in comparing the primarily split in the C&RT analysis compared to the second split on motivation (in Fig. 11 compare the split at nodes 1 and 2 to the split at nodes 25 and 26). The second split was able to identify 8 students that varied from the pattern of lower motivation to read going with more off-task responses.

The third logistic regression looked at the two design features. The only significant predictor of off-task responses was the number lessons that students worked with in ITSS (Wald = 54.72, df = 1, $p = .03$); the further the student worked in the ITSS lessons the higher their off-task responses score ($r = .09$, $p = .24$). The statistical model with this predictor correctly predicted 29.9% of no off-task responses and 83.1% of some off-task responses [Nagelkerke's (1991) $R^2 = .04$]. This suggests that students hurrying through ITSS may have made their rapid progression by submitting off-task responses rather than diligently thinking about the tasks. Slowing down students who tend to submit off-task responses in the system may be a good design change. Such slowing could be accomplished through more individually tailored lessons based on on-line performance as well as more scaffolded practice directed at deficits. In conjunction with more individualized instruction, higher levels of performance could be required before moving to the next text structure.

A logistic regression with the four demographic variables showed gender to be the only significant predictor of no explicit off-task responses versus some off-task responses (Wald = 11.73, df = 1, $p = .001$). The statistical model with gender correctly predicted 65.3% of no off-task responses and 61.6% of some off-task responses [Nagelkerke's (1991) $R^2 = .09$]. Boys submitted more off-task responses than girls (χ^2 1, $N = 186$) = 9.57, $p = .002$).

3.3.3. Using both analysis methods to understand why boys game more than girls

Finally, the significant predictors from the above logistic regression analyses plus other linear predictors identified in the decision-tree analysis were entered in one analysis and stepped into the analysis in the following order: motivational/attitudinal variables (all but performance avoidance that did not appear in the decision-tree analysis), skills and strategy variables (comparison main idea task in ITSS

Table 6
Gain and average off-task responses for 5 terminal nodes predicting off-task behaviors in the comparison lessons (without performance on main idea and recall tasks).

Node	*n*	Gain %	*M* Off-task	Description (motivation to read = MR, main ideas task in ITSS = MIT), performance-approach goals = PA
4	7	3.8%	18.29	Low motivation to read (MR) + high work avoidance
7	8	4.3%	15.00	High MR, lower work avoidance, low attitudes about computers
5	45	24.2%	7.38	High MR + low self-efficacy
8	30	16.1%	6.03	Low MR, lower work avoidance, higher attitudes about computers
6	96	51.6%	2.84	High MR + high self-efficacy

lessons, comparison recall task in ITSS, pretest comparison total recall scores, GSRT reading comprehension test scores), design features (number of lessons and feedback), and then gender. This analysis was used to examine whether or not the other variables could explain differences in off-task responses that could be attributed to gender.

The motivational and computer attitudes variables entered in the first block of variables could account for 11% of the variance (Nagelkerke's (1991) $R^2 = .11$). The skills and strategies variables entered in the second block explained 31% more of the variance (Nagelkerke's $R^2 = .42$). None of the variables in the later blocks could explain additional variance, and thus all of the variance attributed to gender could be explained by the variables in the equation displayed in Table 7. The most powerful predictor in the regression analysis is the average performance of students on the main idea tasks within ITSS. The regression model could explain 68.4% of no explicit off-task responses and 87.1% of some off-task responses. This also was a powerful predictor for off-task responses in Fig. 11 from the decision-tree analysis.

Table 7
Summary of logistic regression with predictor variables for off-task behaviors.

Variables	*B*	SE *B*	Wald	df	*p*
Pretest motivation to read	.008	.038	.046	1	.83
Pretest self-efficacy	−.004	.035	.016	1	.90
Work avoidance	−.011	.043	.671	1	.80
Work mastery	−.015	.059	.065	1	.80
Performance approach	−.006	.060	.010	1	.92
Pretest computer attitudes	−.029	.023	1.534	1	.22
ITSS main idea tasks	−.078	.027	8.454	1	.004
ITSS recall tasks	−.003	.011	.081	1	.78
Pretest total recall	−.003	.019	.030	1	.86
GSRT reading	.021	.026	.664	1	.42
Constant	6.238	2.947	4.48	1	.03

The two analysis approaches can broaden our understanding of why boys submit more explicit off-task responses than girls. This question can be answered in the two complimentary ways depicted in Figs. 13 and 14. Generally boys are doing less work meeting the demand of the ITSS main idea tasks. A third of each gender produces less relevant work to the main idea tasks than the other two-thirds of their gender group (compare nodes 3–6 in Fig. 13). The division shown for each gender in Fig. 13 is between the group scoring nearly a standard deviation or more below the mean versus those scoring higher than this on average for the main idea tasks in ITSS (boys: *M* on main idea task = 45.92, SD = 21.74; girls: *M* on main idea task = 58.49.92, SD = 16.26). However, the bottom third of boys more frequently

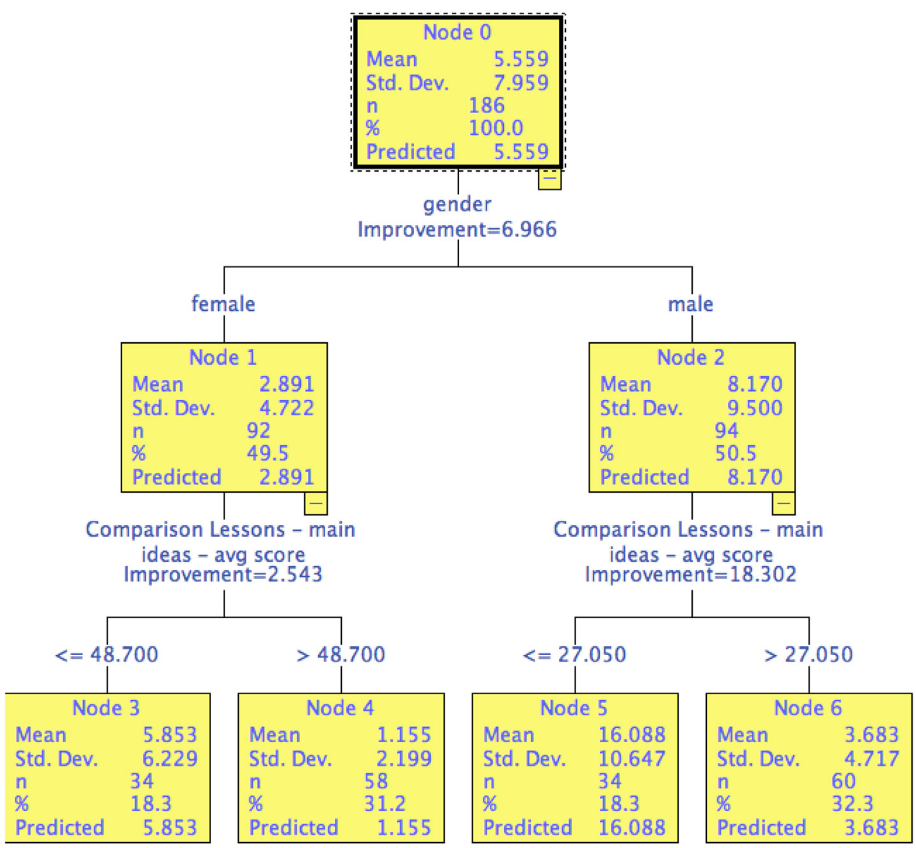

Fig. 13. Why boys produce more off-task responses than girls: ITSS main ideas task performance.

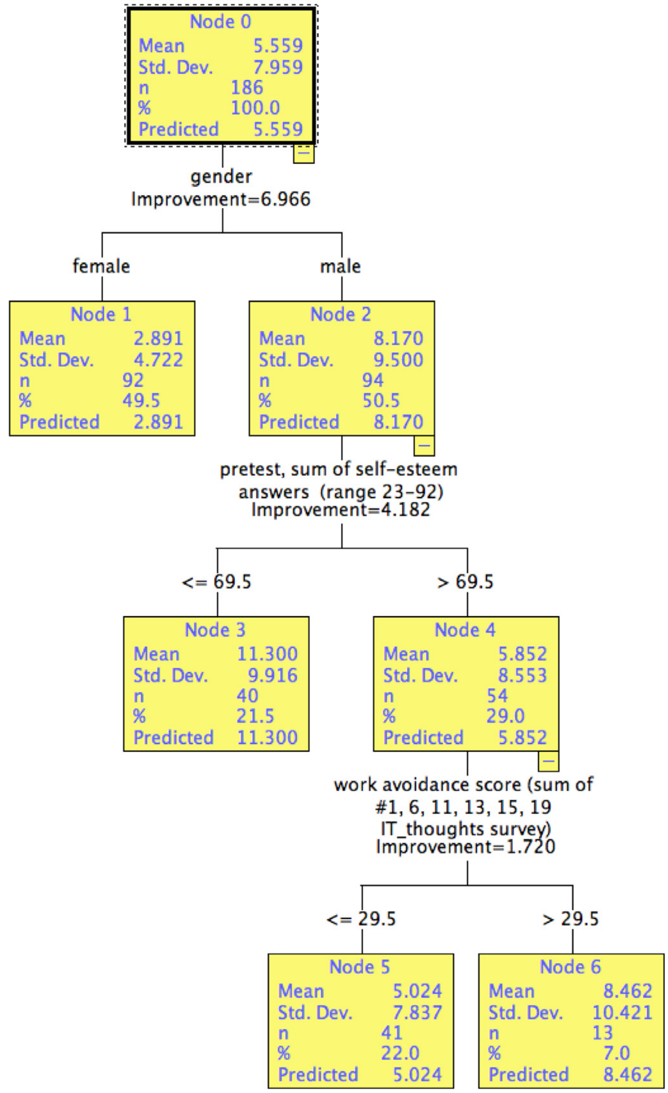

Fig. 14. Why boys produce more off-task responses than girls: low self-efficacy & high work avoidance.

submits explicitly off-task responses than the bottom third of girls. The extreme "off-task responders" are boys and their more frequent off-task responses appears to relate to a) their lack of self-efficacy in terms of anticipating success after committing themselves to working hard on a task and b) their desire to avoid work related to ITSS (see Fig. 14).

4. General discussion

Disengagement is a problem for ITSS as it is with similar computer tutors. Individual differences factors can be helpful in explaining who are the students submitting off-task responses in the system and what safeguards can be introduced into the software to reduce off-task responses and better meet the needs of particular groups of learners. Instruction better matched to a student's skill level or enabled through scaffolding should help to decrease off-task responses for some students (O'Donnell et al., 2007).

4.1. Empirical contribution

This study showed how on-line tracking of average performances during practice trials in learning a cognitive strategy could help to identify students who will game the system. Early identification and intervention with such students may avoid poor learning outcomes in the intelligent tutoring context. Such students may benefit from better individualizing of the instruction. It is of interest to note that students in the more individualized version of ITSS (Meyer et al., 2011) exceeded in learning signaling words (i.e., "in contrast") and general reading comprehension skills over students in standard ITSS. Of particular interest for the issue of off-task responses is that similar increases also were found for mastery learning goals while working in ITSS. Helping to move at least a third of the male participants toward mastery goals and away from work avoidance goals and low self-efficacy would be laudable.

Attitudes (e.g., toward computers for learning) and motivations (e.g., work avoidance goals, interest in the domain) prior to starting the program also predicted off-task responses and could be used to flag some students at risk for submitting off-task responses. For structure strategy instruction, texts related to topics of high interest for specific students with low overall reading comprehension (e.g., ways to increase 4-wheeler speed) could be incorporated into the instruction as well as easier versions of these materials for struggling readers.

4.2. Practical contributions

This research presents important findings relevant to students, classrooms, and schools with regard to off-task behaviors in the setting of reading comprehension instruction using an intelligent tutoring system. Computer-based learning environments have not shown the levels of improvements in learning usually projected when they are initially introduced to schools. One possible cause for the lack of good outcomes may be related to off-task responses, which was the focus of this study.

We have instituted some changes in ITSS to reduce off-task responses, but have not systematically studied the effects of these changes. For example, we have varied number of possible tries that a student receives before a model main idea is offered. Also, we have had the computer flag repeated off-task responses and reported off-task responses to students' teachers on a regular basis. Letting older middle school students know that their teachers will be viewing any profanity or odd remarks appeared to reduce this particular type of off-task response when there was systematic follow-up and consistency.

The decision-tree approach versus regression approach to look at off-task responses illuminated different patterns of subgroups based on student characteristics versus an overall, linear view of factors accounting for most of the off-task responses. The decision-tree approach may be particularly helpful to designers in making software more appropriate for different types of students. It is critical that we find ways to successfully eliminate a disengaged approach when off-task responses will reduce critical strategy, skill, and knowledge acquisition that could be increased by more engaged interactions with a pedagogical agent.

4.3. Theoretical contribution

This study adds to the growing body of research (e.g., Baker et al., 2005, 2010; Cocea, Hershkovitz, & Baker, 2009; Muldner et al., 2010, 2011; Rowe et al., 2009; Sabourin et al., 2013; Walonoski & Heffernan, 2006) about the problem of students' disengaged behaviors with on-line intelligent tutors. The study contributes to this work by looking at elementary and middle school students in the context of reading comprehension instruction. The study points to the importance of a student's motivation for the domain (i.e., reading) prior to any instruction as well as motivation to work during the on-line instruction. Additionally, self-efficacy for succeeding through diligent work as well as other attitudinal and motivational characteristics of the young learner affect off-task behaviors within the system. The study responds to call of investigators (e.g., Muldner et al., 2011) concerned about disengagement in ITS for more research with different populations, contexts, learning domains, and analyses. This study showed the primary importance of student features in off-task responses within the context of intelligent tutoring of the structure strategy in the domain of reading comprehension with 9–12 year olds.

4.4. Limitations of the study and future work

In our study we found little evidence that program features related to explicit off-task responses. Other lesson features within ITSS may contribute to off-task responses, but we only systematically manipulated one design feature, feedback. We did not experimentally look at such factors as redundancy within the program, unclear directions, limited capabilities as far as scoring student answers, or immediate feedback to alert teachers of students' off-task responses. Further research is needed to systematically investigate such issues.

Additionally our study was limited to one particular type of intelligent tutoring program, ITSS, and only two school districts in one state. Thus, the generalizability of our findings will require further investigation across age groups, domains of study, and types of intelligent tutoring programs. Also, replication of our findings with upper elementary and middle school students is needed in other contexts.

References

Ames, C. (1992). Classrooms: goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261–271.
Ames, C., & Archer, J. (1988). Achievement goals in the classroom: students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260–267.
Baker, R. S. J. d. (2007). Is gaming the system state-or-trait? Educational data mining through the multi-contextual application of a validated behavioral model. In *Proceedings of the workshop on data mining for user modeling* (pp. 76–80). Corfu, Greece.
Baker, R., Corbett, A., & Koedinger, K. (2009). Educational software features that encourage and discourage "gaming the system". In *Proceedings of the 14th international conference on artificial intelligence in education (AIED'09)* (pp. 475–482). Brighton, UK.
Baker, R., Corbett, A., Koedinger, K., & Wagner, A. (2004). Off-task behavior in the cognitive tutor classroom: when students "game the system". In *Proceedings of the ACM CHI 2004: Computer–human interaction (CHI'04)* (pp. 383–390). Vienna, Austria.
Baker, R. S. d., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human–Computer Studies, 68*, 223–241. http://dx.doi.org/10.1016/j.ijhcs.2009.12.003.
Baker, R. S., Roll, I., Corbett, A., & Koedinger, K. (2005). Do performance goals lead students to game the system?. In *Proceedings of the 12th international conference on artificial intelligence and education (AIED 2005)* (pp. 57–64). Amsterdam, Netherlands.
Baker, R. S. J. d., & Rossi, L. M. (2013). *Assessing the disengaged behaviors of learners* (Chapter 14). In *Design recommendations for adaptive intelligent tutoring systems learner modeling* (Vol. 1)New York: Columbia University Teacher's College.
Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system." Behavior in interactive learning environments. *Journal of Interactive Learning Research, 19*(2), 185–224.
Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 723–733.
Cocea, M., Hershkovitz, A., & Baker, R. S. J. d. (2009). The impact of off-task and gaming behaviors on learning: immediate or aggregate?. In *Proceedings of the 14th international conference on artificial intelligence in education* (pp. 507–514).
Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. http://dx.doi.org/10.1007/BF02310555.
Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048.
Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*, 218–232.
Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. A. (1996). Assessing motivation to read. *Reading Teacher, 49*, 518–533.
Grimes, J. E. (1975). *The thread of discourse*. The Hague, The Netherlands: Mouton.
Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: necessary and illuminating. *Journal of Educational Psychology, 94*, 638–645. http://dx.doi.org/10.1037//0022-0663.94.3.638.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264.
Krauss, I. K., & Hoyer, W. J. (1984). High technology and the older person: age, sex and experience as moderators of attitudes towards computers. In P. K. Robinson, J. Livingston, & J. E. Birren (Eds.), *Aging and technological advances* (pp. 349–350). New York, NY: Plenum.
Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*, 514–523.
Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Amsterdam: North-Holland.
Meyer, B. J. F. (2003). Text coherence and readability. *Topics in Language Disorders, 23*, 204–224.
Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of the top-level structure in text: key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16*, 72–103.
Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal, 21*, 121–143.
Meyer, B. J. F., Middlemiss, W., Theodorou, Brezinski, K. L., McDougall, J., & Bartlett, B. J. (2002). Effects of structure strategy instruction delivered to fifth-grade children via the internet with and without the aid of older adult tutors. *Journal of Educational Psychology, 94*, 486–519.
Meyer, B. J. F., & Poon, L. W. (1997). Age differences in efficiency of reading comprehension from printed versus computer-displayed text. *Educational Gerontology, 23*, 789–807.
Meyer, B. J. F., & Poon, L. W. (2001). Effects of structure strategy training and signaling on recall of text. *Journal of Educational Psychology, 93*, 141–159.
Meyer, B. J. F., & Wijekumar, K. (2007). A web-based tutoring system for the structure strategy: theoretical background, design, and findings. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 347–375). Mahwah, New Jersey: Lawrence Erlbaum Associates.
Meyer, B. J. F., Wijekumar, K. K., & Lin, Y. (2011). Individualizing a web-based structure strategy intervention for fifth graders' comprehension of nonfiction. *Journal of Educational Psychology, 103*(1), 140–168.
Meyer, B. J. F., Wijekumar, K., Middlemiss, W., Higley, Lei, P. K., Meier, C., et al. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly., 45*(1), 62–92.
Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., et al. (2000). *Manual for the patterns of adaptive learning scales (PALS)*. Ann Arbor, MI: University of Michigan. Retrieved from http://www.umich.edu/~pals/PALS%202000_V12Word97.pdf.
Miller, G. A. (2005). Wordnet [database]. Available 28.02.03 at http://www.cogsci.princeton.edu/~wn/.
Muldner, K., Burleson, W., van de Sande, B., & VanLehn, K. (2010). An analysis of gaming behaviors in an intelligent tutoring system. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent tutoring systems: 10th International conference, ITS 2010* (pp. 184–193). Heidelberg, Germany: Springer. Retrieved from http://www.public.asu.edu/~kvanlehn/Stringent/PDF/Muldner_ITS2010.pdf.
Muldner, K., Burleson, W., Van de Sande, B., & VanLehn, K. (2011). An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. *User Modeling and User-Adapted Interaction, 21*(1–2), 99–135.
Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691–692.
National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards*. Washington, D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers.
O'Donnell, A. M., Reeve, J., & Smith, J. K. (2007). *Educational psychology: Reflection for action*. Hoboken, NJ: John Wiley & Sons, Inc.
Penn Treebank. (2005). [Database]. Retrieved from www.cis.upenn.edu/~treebank/home.html.
Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*, 92–104. http://dx.doi.org/10.1006/ceps.1999.1017.
Ray, M. N., & Meyer, B. J. F. (2011). Individual differences in children's knowledge of expository text structures: a review. *International Electronic Journal of Elementary Education (Special Issue on Reading Comprehension), 4*(1), 67–82.
Rodrigo, M. M. T., Baker, R. S. J.d., D'Mello, S., Gonzalez, M. C. T., Lagud, M. C. V., Lim, S. A. L., et al. (2008). Comparing learners' affect while using an intelligent tutoring system and a simulation problem solving game. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. P. Lajoie (Eds.), *9th International conference on intelligent tutoring systems* (pp. 40–49). Heidelberg, Germany: Spring-Verlag Berlin. http://dx.doi.org/10.1007/978-3-540-69132-7_9.
Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. In *Proceedings of the 13th international conference on artificial intelligence in education* (pp. 195–202). Marina del Ray, CA.
Rowe, J., McQuiggan, S., Robison, J., & Lester, J. (2009). Off-task behavior in narrative centered learning environments. In *Proceedings of the fourteenth international conference on artificial intelligence in education (AIED-09)* (pp. 99–106). Brighton, UK.
Sabourin, J. L., Rowe, J. P., Mott, B. W., & Lester, J. C. (2013). Considering alternate futures to classify off-task behavior as emotion self-regulation: a supervised learning approach. *Journal of Educational Data Mining, 5*(1), 9–38.
Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*, 207–231.
Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: construction and validation. *Psychological Reports, 51*, 663–671.
SPSS. (2001). *AnswerTree 3.0 user's guide*. Chicago, IL: SPSS, Inc.
SPSS. (2007). *Classification trees 16.0*. Chicago, IL: SPSS, Inc.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Walonoski, J. A., & Heffernan, N. T. (2006). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In *Proceedings of the 8th international conference on intelligent tutoring systems (ITS'06)* (pp. 382–391). Jhongli, Taiwan.
Wentzel, K. R. (1999). Social-motivational processes and interpersonal relationships: implications for understanding motivation at school. *Journal of Educational Psychology, 91*, 76–97.
Wiederholt, J. L., & Blalock, G. (2000). *Gray Silent Reading Tests (GSRT)*. Austin, TX: PRO-ED.
Wijekumar, K. K., Meyer, B. J. F., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Journal of Educational Technology Research and Development, 60*, 987–1013. http://dx.doi.org/10.1007/s11423-012-9263-4.
Wijekumar, K. K., Meyer, B. J. F., & Lei, P. (2013). High-fidelity implementation of web-based intelligent tutoring system improves fourth and fifth graders content area reading comprehension. *Computers & Education, 68*, 366–379.
Wixon, M. (2013). *Detecting students who are conducting inquiry without thinking fastidiously (WTF) in the context of microworld learning environments* (Master of Science thesis). Worcester, Massachusetts: Worcester Polytechnic Institute.
Wixon, M., de Baker, R. S. J., Gobert, J. D., Ocumpaugh, J., & Bachmann, M. (2012). WTF? Detecting students who are conducting inquiry without thinking fastidiously. In J. Masthoff, B. Mobasher, M. C. Desmarais, & R. Nkambou (Eds.), *UMAP* (pp. 286–296). Springer, ISBN 978-3-642-31453-7.