To appear in *Computer Science Education* Vol. 00, No. 00, Month 20XX, 1–20

Crib Sheets and Exam Performance in a Data Structures Course

Sally Hamouda^a and Clifford A. Shaffer^{a*}

(Received May 2015)

In this paper we study the relationship between the use of "crib sheets" or "cheat sheets" and performance on in-class exams. Our extensive survey of the existing literature shows that it is not decisive on the questions of when or whether crib sheets actually help students to either perform better on an exam or better learn the material. We report on our own detailed analysis for a body of crib sheets created for the final exam in a junior-level Data Structures and Algorithms course. We wanted to determine whether there is any feature of the crib sheets that correlates to good exam scores. Exam performance was compared against a number of potential indicators for quality in a crib sheet. We have found that students performed significantly better on questions at the comprehension level of Bloom's taxonomy when their crib sheet contained good information on the topic, while performance on questions at higher levels of the taxonomy did not show correlation to crib sheet contents. We have also seen that students at certain levels of performance on the final exam (specifically, medium-to-high performance) did relatively better on certain questions than other students at that performance level when they had good coverage of that question's topic on their crib sheet.

Keywords: crib sheets; cheat sheets; exam performance; data structures and algorithms

1. Introduction

Instructors can adopt a range of interventions to encourage student preparation for exams. One class of interventions involves the materials that students are permitted to bring to an examination. In this paper, we survey the literature on "crib sheets" or "cheat sheets." By this term, we mean a small collection of notes that students are sanctioned by the instructor to create during their exam preparation and bring to the exam session as a reference. We find that the literature is not decisive on the questions of when or whether crib sheets actually help students to either perform better on the exam or better learn the material. We also report on our own detailed examination of a body of crib sheets created for the final exam in a junior-level Data Structures and Algorithms course. We wanted to determine whether there are any behaviors detectable from the crib sheets that correlate to good exam scores.

^{*}Corresponding author. Email: shaffer@vt.edu

2. Literature Review

Examinations can be classified as open-book or closed-book. An open-book exam allows the student to access a body of resources. It is possible that this opportunity to use such materials might lead to less preparation for the exam, or that the student spends excessive time during the exam in looking up the answers to questions (Boniface, 1985; Eilertsen and Valdermo, 2000; Broyles et al., 2005; Heijne-Penninga et al., 2008). Crib sheets are an intermediate option between closed- and open-book exams. It is possible that crib sheets avoid the disadvantages of openbook exams, and that the process of preparing the crib sheet might offer its own pedagogical benefits (Trigwell, 1987; de Raadt, 2012), separate from the benefit of having the crib sheet during the exam.

We categorize previous studies on crib sheets as follows:

- The effect of having the crib sheet on students' exam score (Dickson and Miller, 2005; Dickson and Bauer, 2008; Funk and Dickson, 2011; Hindman, 1980; Visco et al., 2007; Whitley, 1996).
- The effect of having a crib sheet versus an open or closed book test on students' exam score (Boniface, 1985; Erbe, 2007; Gharib et al., 2012).
- The psychological effects of having a crib sheet (Erbe, 2007; Gharib et al., 2012; Joyce et al., 1998; Trigwell, 1987).
- The content of a crib sheet (de Raadt, 2012; Edwards and Loch, 2015; Gharib et al., 2012; Ludorf and Clark, 2014).

We examine the literature on each of these aspects in turn.

2.1. Effects on Exam Score

Two specific aspects of crib sheets have been studied in the literature.

- (1) The effect that *preparing* a crib sheet has on exam score.
- (2) The effect that having the crib sheet's information available has on exam score.

Dickson and Bauer (2008) studied whether the act of preparing a crib sheet leads to better performance on tests. In a lower division Developmental Psychology course, students took four exams during the semester. Students constructed and used crib sheets for the first and the third exams. For the second and fourth exams, the instructor did not instruct the students to make crib sheets, and they were not allowed to use one during these exams. The second and fourth exams are not included in the study. Immediately before taking the first and the third exams, students were given an unannounced pretest without their crib sheets. The first exam's pre-test consisted of 15 randomly chosen questions from the 25 questions on the first exam. The third exam's pre-test consisted of all 15 questions on the third exam. At the end of the semester, students reported their opinions on the impact of making and using a crib sheet on their learning and exam performance, and whether preparing the sheets decreased, increased, or had no effect on their anxiety level.

The first and third pretests were analyzed separately. On the first exam, scores for the questions on the exam (i.e., with the crib sheet) were significantly higher than when those same questions were answered on the pretest (i.e., without the crib sheet). Scores on the third exam were significantly higher than scores on the corresponding pretest (recall that the pretest and the third exam were identical). The authors conclude from these results that students performed significantly better when they used their crib sheets than when they answered the same questions without access to their crib sheets. The student surveys showed that 83% of the students thought that using a crib sheet helped them learn the material and improve their exam performance. 75% of the students reported that making a crib sheet reduced their stress during the exams. Dickson and Baur concluded that the act of preparing the crib sheet does not enhance student learning, because it seems that students depend on the information on the crib sheet to answer the exam questions.

A number of questions remain about this study. The students were told before the pretest that the pretest grade would not be counted in their total course grade. That might have affected student performance on the pretest. The authors indicate that the students performed significantly better in the actual test (where they used their crib sheets) on the questions that were repeated from the pretest. They have not performed an analysis to see if the improved performance is due to simply repeating answering the questions, or whether it was due to the information on the crib sheets.

Visco et al. (2007) investigated the use of crib sheets created by students in a thermodynamics course with 10 students. The goal of the study was to determine best practice for how to prepare a crib sheet. The course had three midterm exams and a comprehensive final exam. Students were interviewed two days after each exam. They were asked about how they thought they did on the exam, how they decided on what to include on the crib sheet, how much they used the sheet, if they think that creating the crib sheet was helpful, if they will change the type of content for the crib sheet next time, and if preparing a crib sheet gives insight on learning or preparation for the exam. Students indicated belief that the construction of the crib sheet helped them to prepare for the exam. However, the authors state (with no evidence given) that this was not shown by student exam scores. The authors concluded that the act of preparing a crib sheet and having it does not necessarily enhance student exam score. We note that the number of students in the study (10) is extremely low, and no empirical experimentation was conducted.

Dickson and Miller (2005) investigated the use of crib sheets in an undergraduate Child and Adolescent Development course. Students completed a survey about their preference for using a crib sheet, how much they think it would enhance their grade, and how it would affect their anxiety. On the second day of the semester, students in one section randomly selected whether they would be permitted to use a crib sheet for the first and third exams, or else for the second and fourth exams. Students in the other section were permitted to use a crib sheet for the second and fourth exams only. In both sections, the percentage of students who actually used crib sheets declined over the course of the semester from a high of 100% on the first exam to around 61.5% and 67.9%, respectively, on the fourth exam. This pattern contrasted with students' strong initial prediction that they would use crib sheets (about 95% predicting so). The students expected that having crib sheets would enhance their performance. However, the exam scores for students who used crib sheets were not better than the scores for students who did not used them. Only 40.5% of students reported that the crib sheet did reduce anxiety, whereas originally about 79.2% expected that it would do so.

Hindman (1980) investigated the use of crib sheets on 52 students enrolled in

two sections of an undergraduate Abnormal Psychology course. There was a test every other week, each covering two textbook chapters. One third of the tests were given with no crib sheets allowed. For the remaining tests, students were asked to prepare and use a crib sheet. On the first day of the class, all the procedure were announced and students were offered the opportunity to transfer to another class. Just before taking each test where students were asked to prepare crib sheets, a member of the class drew a slip of paper to determine if the crib sheets could be used in that test or not for the whole class. When crib sheets were not to be used, they were collected before the test began. On each test, students were asked to indicate if they had made a crib sheet or not.

Students were asked to complete a survey at the end of the semester to rate how helpful were the crib sheets to them in two cases: (1) When they prepared them and were not able to use them and (2) When they prepared and used them. The results showed that the number of students who continued to do the crib sheets declined with time. There was no significant difference on exam scores between the three conditions (crib sheets prepared and not used, crib sheets prepared and used, and no crib sheet). The survey results indicate that students believe crib sheets are moderately helpful. There was no significant difference between how helpful students think the crib sheet is in the two cases (prepared and not used versus prepared and used). The authors interpret that this to mean that students did not consider that using the sheet during the test is more helpful than preparing it.

Funk and Dickson (2011) investigated the use of crib sheets in an Introduction to Personality course. For the first two exams, construction and use of crib sheets was not mentioned. Following the second exam, the instructor taught students how to prepare a crib sheet and provided students with written instructions on crib sheet construction. Students were allowed to use one side of an index card. Students were told that they could use their crib sheet during the third exam. Upon arriving at the exam, students turned in their crib sheets. They then took an unannounced pretest without their crib sheets. The pretest was composed of 20 randomly selected items that were identical to items contained in the larger 50-item exam. After finishing and returning the pretest, students completed the exam with their crib sheets. Students prepared for the fourth exam by making a crib sheet as a study strategy, but they understood that they would not be able to use it during testing. Making and using the crib sheets was optional, but students were encouraged to do it. Upon arriving at the exam, students turned in their crib sheets to the instructor. After finishing the exam, all students returned their exam and completed a post test with their crib sheet. The post-test consisted of 20 randomly selected questions that were identical to those from the fourth exam.

The authors found that the students' scores were significantly lower when they were expecting to use a crib sheet but were not allowed to than when they were not expecting to use the cheat sheet but used it. The authors concluded that the preparation of a crib sheet had a negative impact on learning, which they speculated was because doing so leads to dependence on the content in the sheets. There was no significant difference between scores on the exam taken without the crib sheet and identical items on the post-test taken with the crib sheet when they had not been expecting to use it. The authors interpret this to mean that the students had already studied the material included on the crib sheet, and having the sheet provided no additional benefit.

Whitley (1996) investigated the use of crib sheets in a junior-level Introductory Social Psychology course taught in Spring Semester during two consecutive academic years. In each year, students were enrolled in two sections (one called "morning" and one called "midday"). The students were given three exams. In the second year, students were allowed to bring crib sheets to the three exams. Observation indicated that all students in year two used crib sheets for all exams. A t-test was done between students from the first year (who did not have crib sheets) versus students in the second year (who did have crib sheets). The students in the morning section of year two (who used crib sheets) scored higher than morning section students from year one on the first exam, t = 2.74, p < .05, but not on the second, t = 1.57, or the third, t = -0.01. For the midday section students, there was no significant difference on any of the exams between the two years. The author speculated that differences between students in the morning and midday sections lead to this variation in results. The author observed that in the university where the experiment was held, the students in the morning section are generally the students who enrolled late. They might be considered as procrastinators, which might cause them to need more help on the exam through crib sheets, which serve as a memory aid for them. The author speculated that the reason why the same students did not benefit from the crib sheets in the second and the third exams is that the students were not expecting short answer questions in the first exam, and perhaps the preparation notes he provided after the first exam helped them enhance their performance in the following exams.

Table 1 shows a summary for the literature on the effect that preparing and having a crib sheet has on student performance.

Course	Number	Studied As-	Students	Exam
	of Stu-	pects	think it	score
	\mathbf{dents}		helps	enhanced
Introductory De-	53	Preparation	Not studied	Yes
velopmental Psy-				
chology (Dickson				
and Bauer, 2008)				
Thermodynamics	10	Preparation &	Yes	No
(Visco et al.,		Knowledge		
2007)				
Child and Adoles-	54	Preparation &	Yes	No
cent Development		Knowledge		
(Dickson and				
Miller, 2005)				
Abnormal Psy-	52	Preparation &	Yes	No
chology (Hind-		Knowledge		
man, 1980)				
Introduction to	51	Knowledge	Yes	No
Personality (Funk				
and Dickson,				
2011)				
Introductory	136	Preparation &	Not studied	1 in 6 ex-
Social Psychology		Knowledge		ams
(Whitley, 1996)		_		

 Table 1. Summary for the literature on crib sheets and student performance.

In summary, while not all studies agree on whether crib sheets affect student

performance, the weight of the studies appears to argue against this effect. One study showed that they have a positive impact on student performance (Dickson and Bauer, 2008) (with some methodological questions to the study), one showed an effect on one of six exams, and four studies showed no effect (Whitley, 1996).

2.2. Crib Sheets, Open-Book, and Closed-Book Exams

Boniface (1985) studied the use of textbooks and notes during an open-book exam. The experiment was done with 30 volunteers from students in a Psychology Design and Analysis course. The students took a three-hour open-book exam. Each student was observed every five minutes to see if he was consulting a book, consulting notes, writing, or doing something else. At the end of the exam, students were asked to complete a survey asking about their preparation for the exam and their attitude during the exam. The authors made use of the same students' previous scores on two closed-book exams addressing the same subject area to evaluate how strong each student was in the subject area.

The results showed that the amount of time devoted to consulting notes and texts was negatively correlated to the score on the examination. In addition, students who had a history of success in the subject area did not need to use their notes as much as the weaker students. The same result was found by Burns (2014), who saw a negative correlation between the number of times a student used a crib sheet and exam performance. High-achieving students (those with good scores in previous exams on the same area) did not refer to their crib sheets as often.

Erbe (2007) compared three approaches: providing students with an instructorcreated formula sheet, open-book exams, and student-created crib sheets. No empirical studies were done on the three treatments. From the author's observations, he suggests that student-created crib sheets can reduce anxiety while enhancing learning in the courses that assess on the first three levels of Bloom's taxonomy (i.e. knowledge, comprehension and application). Erbe hypothesizes that open-book exams can lead to poor preparation for the exam. He encouraged his students to build a good crib sheet by awarding a prize for the most creative one. He did not allow students to use photo-copies of books or articles. He observed that students rarely needed the crib sheet they prepared, however he states without evidence that they "loved" the idea of having a crib sheet for exams. He hypothesizes with no evidence that the act of preparing the crib sheet leads to better learning than having access to an instructor-provided formula sheet.

Gharib et al. (2012) studied 297 students enrolled in eight different sections of Introductory Psychology and 99 students enrolled in four sections of a Statistics course. The authors measured test anxiety in the two courses.

In the Introductory Psychology course, all sections were taught by the same instructor and had the same texts and assignments. During a given term, identical exams (non-cumulative, 50 multiple-choice questions on each exam) were given to all sections. The eight sections of Introductory Psychology involved in the study were taught over four terms. Student exam scores on closed-book exams were compared to exam scores on crib sheet and open-book tests in three sections, all taught the same term. The three exam treatments were counterbalanced among the three sections, so that each exam was given in all the three formats (to different sections). In the remaining five sections, the first two exams were either open-book or crib sheet, counterbalanced across four sections.

In the Statistics course, all sections were taught by the same instructor and had

the same textbook, homework assignments, and exams. There were two sections taught each term for two terms. There were two exams, and the exams were a combination of short answer and story problems. The two exams were counterbalanced among the four sections for the two treatments of open-book and crib sheet.

After the first exam in each section of class (both in Introductory Psychology and Statistics), the students completed the exam preference questionnaire, and also reported the number of hours that they had spent on studying for that exam. Students indicated a preference for open-book and crib sheet exams over closedbook exams.

In Introductory Psychology, paired-samples t-tests revealed that closed-book exams resulted in lower scores than either open-book exams or crib sheet exams, and crib sheet exams resulted in lower scores than open-book exams. In Statistics, the difference in exam scores between open-book and crib sheet exams was not significant. Pearson correlations reveal that scores on the different exam treatments are positively correlated in both classes for all the treatment types. In other words, students who do well on one exam treatment do well on the other two types. It was not studied whether students actually learn more under one of these treatments (as opposed to perhaps doing better on the exam with more information available simply because of the information).

For both courses, a Pearson correlation shows that the test anxiety measured right before the exam was negatively correlated with scores on the exam. A pairedsamples t-test finds higher anxiety scores during exams using crib sheets compared to the open-book exam in both Introductory Psychology and Statistics. Test anxiety was not measured for the closed-book treatment. The results of the student preferences questionnaire showed that both classes predicted that they would do better on open-book or crib sheet exams compared to closed-book exams. Comparisons of the actual exam scores of those who predicted they would do best on open-book exams and those who predicted they would perform best on crib sheet exam were made. An independent samples t-test finds that preference for openbook exams versus crib sheets did not predict relative performance for the two treatments.

2.3. Psychological Effects

Joyce et al. (1998) examined student-developed crib sheets created for a Psychiatric/Mental Health nursing course. The authors observed that 80% of the students brought crib sheets to the exam. Students who brought a crib sheet were asked to provide written feedback regarding the effectiveness of the use of the crib sheet on the exam. The authors interpret those written responses to indicate that the crib sheets reduced anxiety.

Trigwell (1987) support the same hypothesis on the use of crib sheets, based on the comments gathered from student interviews in a first-year science course. Students indicated that having the crib sheet at the exam made them feel more secure. Erbe (2007) found the same results in courses including Statistics, Research Methods, Methods for Teaching Mathematics, and Computer Use in Education. However, he does not show empirical evidence or formal studies.

As mentioned previously, Gharib et al. (2012) also studied physiological effects. Their experiments involved five Introductory Psychology sections and two Statistics sections. The students completed the Pre-Examination Worry Emotionality Scale (Morris et al., 1981), a 10-item measure of test anxiety that asks (on a fivepoint likert scale) questions about current levels of negative emotionality ("I feel my heart beating fast") and worry ("I am afraid that I should have studied more for this test"). The overall scores were used as a measure of test anxiety.

For both courses, a Pearson correlation shows that the test anxiety measured right before the exam was negatively correlated with scores on the exam. A pairedsamples t-test finds higher anxiety scores during crib sheet exams compared to open-book exams in both Introductory Psychology and Statistics.

As mentioned in Section 2.1, the last survey offered by Dickson and Miller (2005) indicated that the majority of the students do not believe that crib sheets reduce anxiety.

Dickson and Bauer (2008) tried to determine if having a crib sheet has an effect on students' anxiety. At the end of the semester, students reported the impact of making and using a crib sheet on their learning and exam performance, and whether it decreased, increased, or had no effect on their anxiety level. The results of the survey showed that about 75% of the students reported that making a crib sheet reduced their stress during exams.

2.4. Crib Sheet Content

Recent studies (Ludorf and Clark, 2014; de Raadt, 2012; Gharib et al., 2012) recommended more investigation on the effects of the content of the crib sheets in different disciplines. Most crib sheet studies have been done on courses related to psychology and statistics. One study was done on a Thermodynamics course (Visco et al., 2007) and another on an Introduction to Programming course (de Raadt, 2012).

de Raadt (2012) reported on the use of crib sheets in an introductory programming course. In addition to their crib sheet, students were given a programming language reference during the exam. The students had previously seen this reference guide during a sample exam. Students were asked not to include such information in their crib sheets. About 20% of the students choose not to bring a crib sheet to the exam. A t-test found that the group who brought a crib sheet had significantly better scores than the group who did not. de Raadt categorized the crib sheets based on layout features and content features. The layout features studied were density, organization, and whether the content order matched that of the course content. The content features were code examples, abstract representations, sample answers, and language reference. Each of the features was identified as either present or absent. The author found that the following features were positively correlated to student exam score.

- (1) The ordering of the crib sheet to match the course content. The authors interpret that it could be an indication of a thorough review of the course content.
- (2) The organization of information. This could mean that it was easier for the students to find the information they need.
- (3) Inclusion of abstract representations of content, meaning whether students represented the concepts in a general way, using text or diagrams, rather than program code for specific examples. The authors interpreted that this could relate to the adaptability to abstract content to new problems.

Gharib et al. (2012) also studied the relationship between the quality of crib sheets and exam performance. Student-prepared crib sheets were rated on a 10-

point scale based on the organization of the crib sheet (whether there are headings and subheadings, highlights, etc.) and amount of detail (number of words). Rating was done by a rater blind to the students' exam scores. Each crib sheet was assigned a single score to represent its quality. A Pearson correlation found that the quality of the crib sheet was not related to scores on Introductory Psychology exams, but there was a positive correlation between crib sheet quality and exam score in Statistics.

Edwards and Loch (2015) looked at the the content and layout of crib sheets used by students for an end-of-semester calculus exam. They sought to determine features that best characterize crib sheets. They studied layout-based features, including density, emphasis, and sheet structure. Content-based features included examples, representations, formulae, meta-content (like reminders, messages or arrows), correctness (whether the sheet had errors), and completeness. All of the features were binary indicators, indicating whether the crib sheet had the feature or not. The authors did not compute correlations between those features and students exam scores. The authors state that their initial analysis addresses a gap in the literature regarding how crib sheets can be characterized based on their contents.

Ludorf and Clark (2014) evaluated crib sheets from students enrolled in a juniorlevel Psychological Statistics course. Crib sheets were evaluated on overall quality, verbal process information, numeric process information, organization of information, use of color, and submission order (ordinal position when the test was submitted). They found that higher quality crib sheets correlated to higher test scores. They found also that higher test scores correlated to lower density of information.

Higher verbal process rating, which means that the sheet included information like instructions on how to perform some process, were associated with lower test scores. None of the other variables were correlated to performance. The authors hypothesize that crib sheets with low information density are created by students with better understanding, so they don't feel the need to include a lot of information.

Table 2 shows a summary of the features used in the literature to characterize crib sheets.

3. Methods

In this section, we present results from a detailed study that we made on the crib sheets created by students for the final exam in a junior-level computer science course on Data Structures and Algorithms. We first examined the crib sheets to identify distinctive features that seem like plausible candidates for having an effect on exam scores. We then performed statistical tests to determine whether any of these features are in fact correlated to exam scores. Our research questions are as follows.

- Do any of the crib sheet distinctive features has correlation to exam scores?
- Does including a specific topic in the crib sheet lead to a better performance on the questions covering this topic?
- Do students with similar performance on the exam produce crib sheets with similar quality?
- Do questions at certain levels of Blooms taxonomy benefit from the crib sheet

Course	Number	Features	Most effective
	of crib		features
	sheets		
Introduction	89	Density, organization, or-	Organization,
to Program-		dering, examples, abstract	ordering, abstract
ming (de Raadt,		content, sample answers,	content
2012)		language reference duplica-	
		tion	
Statistics (Gharib	99	Organization, amount of	Not studied
et al., 2012)		detail	
Physiological	21	Overall quality, verbal pro-	Overall quality
Statistics (Ludorf		cess information, numeric	and density
and Clark, 2014)		process information, den-	
		sity of information, organi-	
		zation of information, use	
		of color, submission order	
Calculus (Ed-	30	Density, emphasis, sheet	Not Studied
wards and Loch,		structure, examples, repre-	
2015)		sentations, formulae, meta-	
		content, correctness	

 Table 2. A summary of literature on crib sheet content

content?

• Do students at certain levels of performance benefit more from their crib sheet content?

We examined the relationship between crib sheets and student performance in two sections of CS3114 Data Structures and Algorithm Analysis at Virginia Tech, involving a total of approximately 150 students. This course is taken after a standard CS2 course, typically by second semester sophomores or first semester juniors. Upon completion of the course, students are expected to have mastered the following skills:

- Choose the data structures that effectively model the information in a problem.
- Judge efficiency trade-offs among alternative data structure implementations or combinations.
- Apply algorithm analysis techniques to evaluate the performance of an algorithm and to compare data structures.
- Implement and know when to apply standard algorithms for searching and sorting.
- Design, implement, test, and debug programs using a variety of data structures including buffer pools, hash tables, and advanced tree structures.
- Select appropriate methods for organizing data files and implement file-based data structures.

A typical exam for CS3114 tries to test students on the items listed above. While there is some content to recall, the exam does not heavily depend on memorization or applying certain formulas. Students do need to know specific information about a particular collection of data structures and algorithms, including their performance (which many students consider to be a memorization task). Many of the test questions ask students to explain an algorithmic process, use information to solve a calculation, or choose the best solution between a set of design alternatives. More details about the exam are presented in Section 5.

Students were permitted to create and bring with them one $8 \ 1/2 \times 11$ inch sheet of paper with notes for the final exam. The students were used to this procedure for crib sheets, as they had been allowed similar crib sheets on their midterms. The crib sheet could contain any information a student wanted to include, and both sides of the sheet could be used. This exam counted for 15% of the semester grade (and all raw scores that we report in this section are out of 150 points).

At the end of the final, the crib sheets were collected. The students were unaware that an assessment would be done on their crib sheets when they prepared them. Students were given the opportunity to sign informed consent at the beginning of the semester to permit use of their materials and performance for research purposes. Only crib sheets of students who signed the consent form are used in this study. While the crib sheets themselves contained identification so that they could be related to exam performance, in no way did the existence or quality of the crib sheet affect the grading process, as analysis of the crib sheets was done well after the course was completed.

The first step in our process was to examine the crib sheets to determine distinctive features that we hypothesized might be relevant to exam performance. Then, after deciding on those distinctive features, the crib sheets were evaluated on the selected features shown in Table 3. During the evaluation step, the rater was blind to students' scores on the exam.

Variable	Description	Possible
		Values
Overall Quality	An overall quality score assigned by a	1-5 (5 is the
	rater to the crib sheet	best)
Density	Percentage of the available space that	0-100%
	was used	
Organization	Subjective assessment by the rater for	1-5 (5 is the
	how well the crib sheet is organized	best)
Definitions	The percentage of information in the	0-100%
	form of definitions	
Printed or Written	Whether the crib sheet was printed,	p/w/mix
	written, or mixed	
Main Topics	The topics of interest	varies
Covered Material	Percentage of course material that is	0-100%
	covered	
Copied Material	Percentage of information that is di-	0-100%
	rectly copied from course material	
Code Examples	Percentage of content as code examples	0-100%
Use Color	Uses color or not	y/n

Table 3. Distinctive features found in our crib sheets

We used some features discussed in the literature, including overall quality, density, organization and color use. We added features that we think are distinctive to our crib sheets, including printed or written, main topics, covered material, copied materials, and code examples.

To find the features of a given crib sheet, the rater iterated through all the sheets

to make a thorough manual inspection. In detail, the crib sheets were evaluated as follows.

- (1) **Overall Quality** is a subjective measure based on overall examination, measured by a value from 1 to 5. A crib sheet with quality 5 appears to be well organized, has a variety of information covering most of the course topics, and it is easy to spot information in the sheet without taking much time (such as if the information is organized in a table). Examples of high quality and low quality sheets are provided in Appendix B. Sheets with similar quality from the rater's point of view were put in a pile. Initially, the rater had 5 empty piles numbered 1 through 5. The rater first iterated over the sheets to get a rough opinion on characteristics of sheets that belong to each quality level. Then, the rater looked at each sheet in turn to assign it to a quality pile. At the end of the rating process, each pile had a set of sheets that looked similar in terms of quality. The rater iterated then over each pile to make sure that all of the sheets in a given pile have similar quality. If the rater found a sheet that looked out of place, then the sheet was moved.
- (2) **Density** is the percentage of the available space that is used by the student, measured to the nearest 10%. Student were permitted to use both sides of the 8 $1/2 \times 11$ inch sheet. A sheet where the student used all of the space on both sides is given a density value of 100%. If one side is used then the value given to the density is 50%. None of the sheets used less than 10% of the space available. Density indicates how much the information covered by the sheet, however, it does not indicate how many topics are covered by the sheet. If the sheet is dense, it could be because the student covered a certain topic in great detail.
- (3) **Organization** is a subjective assessment for how well the sheet is organized, on a scale of 1 to 5. A process similar to that used to assess Overall Quality was used to determine the Organization score. Sheets with best organization had headers and sub-headers, and tables for comparisons making them easy to ready and spot information.
- (4) **Definitions** is a measure for what fraction of what is written on the sheet is definitions, to the nearest 10%. The rater looked at each sheet to see how much of the written material was purely definitions in the form of "term: definition". For this exam, pure definitions tended to be of limited help because the questions were mainly focused on understanding.
- (5) **Printed or Written** is a categorical indicator for whether the sheet content is (1) printed, (2) written, or (3) a mix of both. This feature may indicate how much effort a student exerted in preparing the sheet. Written sheets could indicate much effort exerted in preparing the sheet. Printed sheets might indicate text copied and pasted from course materials, or might indicate material was well prepared on a computer. Printed materials tend to easy to read (though not necessarily well organized).
- (6) **Main topics** is a list of the topics covered by the sheet. The rater used a model list of topics covered by the exam. For each sheet, the rater looked to see which topics from this list are covered by the sheet. For each sheet, each topic was given a value of 0 if it is not covered, and a 1 if it is covered. If topics covered by the sheet matches the topics covered by the exam, then the sheet could be of help to the student.
- (7) Covered Material is the percentage of the course topics that are covered

by the sheet, measured to the nearest 10%. For each sheet, the rater took the fraction of topics covered by the sheet compared to the number of topics covered by the course. Higher topic coverage could be an indicator that the student understands the course content, and knows how to relate different course topics to each other.

- (8) **Copied Material** is the percentage of the sheet content that is copied from course materials with no interpretation, rounded to the nearest 10%. For each sheet, the rater estimated the percentage of the sheet content that is copied from the course online material. We hypothesize that the more the directly copied material the less effort the student put to the sheet, the less understanding the student has for the material, and the less the synthesis the student has done to compile the course material.
- (9) Code Examples is the percentage of the sheet content that is code examples, rounded to the nearest 10%. For each sheet, the rater estimated the percentage of the sheet content that is code examples or pseudo code.
- (10) **Use Color** is a binary indicator for whether the student used colors on the sheet or not. Any sheet which uses any color (beyond the sheet background and text foreground) is considered to be using color. Color use could lead to a more readable sheet, and might indicate greater care in preparation.

4. Results

The analysis presented here is based on students' crib sheets and test scores for a single test, the course final exam. We calculated the correlation between each of the variables of interest described in Table 3 and the students' final score in order to determine if we can find a relationship. We did not find any such correlation for the overall exam, but see the item-level analysis presented in Section 5.

With and without crib sheet: 104 students brought a crib sheet to the exam, and 27 did not. A t-test showed no significant difference in exam score between the two groups (t=0.1744, p=0.8618). Table 4 shows the t-test result.

	With	Without
Mean	110.29	109.44
Standard Deviation	21.87	24.4
Standard error of the mean	2.14	4.7
Number of Students	104	27

Table 4. t-test comparing with and without crib sheet groups

With and without color: 24 students used colors on their crib sheet versus 80 who did not. A t-test showed no significant difference between the two groups with respect to exam score (t=1.1502, p=0.2527). Table 5 shows the t-test result.

	Uncolored	Colored
Mean	111.64	105.79
Standard Deviation	20.99	24.53
Standard error of the mean	2.35	5.01
Number of Students	80	24

Table 5. t-test for colored versus uncolored crib sheets

Printed or Written: 52 students used a printed crib sheet, 47 used a handwritten sheet, and five had a mix. A t-test showed no significant difference between the group who had a printed crib sheets versus the group who had a hand-written crib sheets (t=0.3314, p= 0.7410). Table 6 shows the t-test result.

	Printed	Written
Mean	110.87	109.38
Standard Deviation	22.33	22.11
Standard error of the mean	3.10	3.23
Number of Students	52	47

Topic-level analysis: For each crib sheet, we categorized the topics included. Almost every question on the exam relates to a specific topic. For each topic covered in the exam, we have done a t-test to determine if the students who included the topic in their crib sheets scored better or not. For all topics, the t-tests showed no significant difference with respect to the overall exam. However, in Section 5 we analyze the relationship between specific topics and specific exam questions.

Shared crib sheets: We found four groups of students where the members of a group had the same crib sheet. We assume that each group either jointly created the sheet, or one person shared theirs with the other members of the group. All of the shared crib sheets were printed rather than hand written. We have noticed that within each group there is a substantial range in scores, as can be seen in Table 7.

Table 7.	Groups	with	similar	crib	sheets.
----------	--------	------	---------	-----------------------	---------

Group	Member	Score
Group 1	1	105
	2	115
	3	57
Group 2	1	99
	2	131
Group 3	1	102
	2	114
Group 4	1	104
	2	117

Exam score and crib sheet overall quality: In an attempt to examine whether student proficiency with the course content relates to the crib sheets' overall quality, we have clustered the students into quartiles based on their final exam score. Then we performed a t-test among all of the quartiles. For each pair of quartiles, we test if there is a significant difference in the crib sheets' overall quality. None of the t-tests showed any significant difference. The results showed that the first and fourth quartiles had the largest mean difference in overall crib sheet quality. Details are shown in Table 8.

We have also checked in the other direction, to see if stratification by crib sheets' overall quality reveals a difference in group exam scores. We clustered the students into quartiles based on their crib sheets overall quality and did a t-test among all the quartile pairs. None of the t-tests showed any significant difference.

	First	Fourth
Mean	1.82	2.2
Standard Deviation	1.05	1.04
Standard error of the mean	0.21	0.21
Number of Students	25	25

Table 8. Crib sheet overall quality: t-test of the first versus the fourth score quartile

5. Exam Analysis

We did several analyses on the exam itself. It is possible that characteristics of an exam relate to the effectiveness of crib sheets. It is possible that exam characteristics explain some of the difference in effectiveness outcomes reported in the literature. Most research papers on the subject do not provide details about the exam used.

Appendix A provides the difficulty, discrimination indices, and item characteristics curves for the exam. We have used the ltm¹ R package to perform item analysis on the test questions. We used the two-parameter logistic model, which takes into consideration the discrimination and the difficulty of a question. The IRCs demonstrate the desired correlation between conceptual knowledge and item performance for all exam questions. As student ability increases, the probability to solve the question correctly increases as well and vice versa.

The IRCs show a range of difficulties on the questions. For two of the questions, a student of average ability is less than 50% likely to get the question correct. Two or three of the questions appear to be relatively easy for a student of medium ability, and two or three of the questions are in between. This seems to imply a good range of difficulties, and the median score on the exam is typically around 70% or so. Table A1 shows the difficulty index, relates to the point on the x-axis where the 50% probability of getting the question correct occurs. A negative value indicates an easier question than a positive value. The discrimination index indicates the slope of the IRC curve. The higher the discrimination index, the more sharply it discriminates between students with slightly lower or higher ability. Alternatively, it indicates the difference between questions that tend to be completely right or completely wrong, versus those more likely to yield partial credit.

5.1. Bloom's Taxonomy Level

The final exam had seven questions. The Bloom's taxonomy level Forehand (2010) for each question is as follows.

- (1) Question 1: Compare space requirements of adjacency matrix and adjacency list. This question is at the comprehension level.
- (2) Question 2: Comparison of implementations for Dijkstra's algorithm. This question is at the comprehension level.
- (3) Question 3: Proficiency exercise on Union/Find. This question is at the application level.
- (4) Question 4: Write a recursive function (BST range query). This question is at the application level.

 $^{^{1} \}rm https://cran.r-project.org/web/packages/ltm/$

- (5) Question 5: File Processing questions. These questions are at the comprehension level.
- (6) Question 6: Proficiency exercise on B⁺ trees. This question is at the application level.
- (7) Question 7: Reasons to use a B^+ tree instead of a BST. This question is at the comprehension level.

We are interested in testing the hypothesis that the Bloom's taxonomy level relates to crib sheet effectiveness. We examined the crib sheets and give a binary value of 0 or 1 for whether the crib sheet contains useful information related to each question of the exam. A crib sheet gets a mark of 1 for a certain topic if it well covers that question and 0 otherwise. Then for each question, we have performed a t-test to see if students who included the topic related to the question in their crib sheets scored better than the students who did not.

Tables 9, 10, 11, and 12 shows a significant difference between the scores of the students who included the topic versus the students who did not include it in their crib sheets. Students who included the topic related to the corresponding question in their crib sheet scored better than who did not. The p values are 0.0265, 0.0101, 0.0409, and 0.0046 for questions 1, 2, 5, and 7 respectively. T-tests for the other questions did not show any significant difference. We can see that Questions 1, 2, 5, and 7 are all at the comprehension level of Bloom's taxonomy, while the other questions are at higher levels. Questions at levels higher than the comprehension level are hard to answer directly from the crib sheet, as they require a deeper understanding beyond pulling information from the crib sheet.

Table 9. Question 1: t-test of the sheets which included the corresponding topic versus the sheets which did not

	Not included	Included
Mean	10.85	13.34
Standard Deviation	6.62	4.21
Standard error of the mean	1.27	0.48
Number of Students	27	77

 Table 10. Question 2: t-test of the sheets which included the corresponding topic versus the sheets which did not

	Not included	Included
Mean	5.61	8.93
Standard Deviation	6.59	6.17
Standard error of the mean	0.86	0.92
Number of Students	59	45

Table 11. Question 5: t-test of the sheets which included the corresponding topic versus the sheets which did not

	Not included	Included
Mean	9.24	11.24
Standard Deviation	4.29	5.52
Standard error of the mean	0.55	0.85
Number of Students	62	42

	Not included	Included
Mean	15.29	18.27
Standard Deviation	6.02	4.29
Standard error of the mean	1.03	0.51
Number of Students	34	70

 Table 12. Question 7: t-test of the sheets which included the corresponding topic versus the sheets which did not

5.2. Student overall performance

We have grouped students into groups based on their overall final exam score. The scores of the students range from 37 to 150.

We have clustered student who have close scores (within 20 points). So we have a group for students who scored from 150 to 130, another from 129 to 110, and so on. For each group we looked at each question on the comprehension level to see if the quality of the sheets with better coverage for the topic covered by the question correlated to a better score on that question.

Table 13 shows that for Question 2, students in the group that has scores ranging from 90 to 110 had statistically significant better grades when they included the topic of the question on their crib sheets. The p-value was 0.0241.

Table 14 shows that for question 5, students in the group that has scores ranging from 130 to 150 had statistically significant better grades when they included the topic of the question on their crib sheets. The p-value was 0.0288.

Table 15 shows that for question 7, students in group that has scores ranging from 90 to 110 had statistically significant better grades when they included the topic of the question on their crib sheets. The p-value was 0.0013.

We can see that students who are considered to be in the medium or high performance level sometimes benefit from what they have included in their crib sheets for some of the comprehension questions, if the topic related to the question is well covered in the crib sheet. We did not find any significant difference for other pairings of questions and performance groups. Given that there were six groupings of student by performance, and four questions at the comprehension level, this means that we examined 24 separate pairings. Of those pairings, only three suggest a causal relationship between good coverage of the associated topic on the crib sheet and question performance. We conclude that while preparing complete information on a crib sheet is indeed likely to improve test scores for questions low on the Bloom's taxonomy, this effect is likely to be weak. Furthermore, this does not in itself address the question of whether preparing crib sheets aids learning (as separate from short-term test performance).

	Not included	Included
Mean	2.89	7.6
Standard Deviation	5.22	6.38
Standard error of the mean	1.2	1.65
Number of Students	19	15

Table 13. Question 2: t-test of the sheets which included the corresponding topic versus the sheets which did not for students within the range of final exam scores from 90 to 110

	Not included	Included
Mean	13.1	18.3
Standard Deviation	6.4	2.63
Standard error of the mean	2.02	0.83
Number of Students	10	10

Table 14. Question 5: t-test of the sheets which included the corresponding topic versus the sheets which did not for students within the range of final exam scores from 130 to 150

Table 15. Question 7: t-test of the sheets which included the corresponding topic versus the sheets which did not for students within the range of final exam scores from 90 to 110

	Not included	Included
Mean	11.36	17.83
Standard Deviation	5.95	4.48
Standard error of the mean	1.8	0.93
Number of Students	11	23

6. Conclusions

We performed a thorough review of the literature related to crib sheets and exam performance. We also conducted our own empirical study of the relationship between crib sheet features versus exam scores. We have reached the following conclusions.

- There is little support in the literature for the hypothesis that preparing crib sheets improves exam performance.
- While the limited research on relative performance between open- and closedbook exams indicates that students score higher on the same exam when given in open-book form, there is no research that we are aware of to indicate that this translates to better learning of the material (as opposed to simply being able to answer some additional questions from content in the notes).
- Students routinely indicate that they prefer exams that are either open-book or allow crib sheets.
- Empirical evaluations of anxiety indicate that use of open-book exams or exams with crib sheets reduce anxiety.

Based on detailed analysis of the exam used for this study, we found that students performed better on questions at the comprehension level of Bloom's taxonomy when their crib sheet contained good information on the topic, while performance on questions at higher levels of the taxonomy did not show correlation to crib sheet contents. When grouped by performance level, we found that among students with medium performance in the final exam, those with crib sheet content that covered a specific question (at the comprehension level) often performed better on that question.

7. Acknowledgements

This work is supported in part by the National Science Foundation, under grants DUE-1139861 and IIS-1258471.

References

- David Boniface. Candidates' use of notes and textbooks during an open-book examination. *Educational Research*, 27(3):201–209, 1985.
- India L Broyles, Peggy R Cyr, and Neil Korsen. Open book tests: Assessment of academic learning in clerkships. *Medical teacher*, 27(5):456–462, 2005.
- Kathleen C Burns. Security blanket or crutch? Crib card usage depends on students' abilities. *Teaching of Psychology*, 41(1):66–68, 2014.
- Michael de Raadt. Student created cheat-sheets in examinations: Impact on student outcomes. In Proceedings of the Fourteenth Australasian Computing Education Conference-Volume 123, pages 71–76. Australian Computer Society, Inc., 2012.
- K Laurie Dickson and Jack J Bauer. Do students learn course material during crib sheet construction? *Teaching of Psychology*, 35(2):117–120, 2008.
- K Laurie Dickson and Michelle D Miller. Authorized crib cards do not improve exam performance. *Teaching of Psychology*, 32(4):230–233, 2005.
- Antony Edwards and Birgit Loch. A preliminary categorization of what mathematics undergraduate students include on exam "crib sheets". In *Conference on* research in undergraduate mathematics education, pages 47–51, 2015.
- Tor Vidar Eilertsen and Odd Valdermo. Open-book assessment: A contribution to improved learning? *Studies in Educational Evaluation*, 26(2):91–103, 2000.
- Brigitte Erbe. Reducing test anxiety while increasing learning: The cheat sheet. College teaching, 55(3):96–98, 2007.
- Mary Forehand. Blooms taxonomy. *Emerging perspectives on learning, teaching, and technology*, pages 41–47, 2010.
- Steven C Funk and K Laurie Dickson. Crib card use during tests helpful or a crutch? *Teaching of Psychology*, 38(2):114–117, 2011.
- Afshin Gharib, William Phillips, and Noelle Mathew. Cheat sheet or open-book? A comparison of the effects of exam types on performance, retention, and anxiety. *Online Submission*, 2(8):469–478, 2012.
- Marjolein Heijne-Penninga, Jan Kuks, WH Hofman, and Janke Cohen-Schotanus. Influence of open-and closed-book tests on medical students' learning approaches. *Medical education*, 42(10):967–974, 2008.
- C Douglas Hindman. Crib notes in the classroom: Cheaters never win. *Teaching* of Psychology, 7(3):166–168, 1980.
- Virginia K Joyce, Patricia Freed Drake, and M Hunter. Crib sheets or security blankets? *Issues in mental health nursing*, 19(3):291–300, 1998.
- Mark R Ludorf and Sarah O Clark. Help sheet content predicts test performance. Essays from E-xcellence in Teaching Volume XIII, page 60, 2014.
- Larry W Morris, Mark A Davis, and Calvin H Hutchings. Cognitive and emotional components of anxiety: Literature review and a revised worry–emotionality scale. *Journal of educational psychology*, 73(4):541, 1981.
- Keith Trigwell. The crib card examination system. Assessment and Evaluation in Higher Education, 12(1):56–65, 1987.
- Donald Visco, S Swaminathan, L Zagumny, and H Anthony. Interpreting studentconstructed study guides. In ASEE Annual Meeting and Exposition Proceedings, pages 1–10, 2007.
- BE Whitley. Does "cheating" help? The effect of using authorized crib notes during examinations. *College Student Journal*, 30:489–493, 1996.

Appendix A: Final Exam Item Response Analysis

Question	Difficulty Index	Discrimination Index
Q1	-1.07	1.75
Q2	0.39	1.71
Q3	-0.84	13.79
Q4	-0.48	1.34
Q5	0.43	1.24
Q6	-1.25	6.01
Q7	-1.42	2.18

Table A1. Difficulty and discrimination indices computed by ltm package



Item Characteristic Curves

Figure A1. Item response curves for the final exam questions

Appendix B: Examples of Crib Sheets

Examples of good, medium, and poor quality crib sheets.



(a) Sheet Front



(b) Sheet Back

Figure B1. An example of a good quality crib sheet.



(a) Sheet Front



(b) Sheet Back

Figure B2. An example of a medium quality crib sheet.

0.	
41.	in 5 new closest to free
01	and these totind maxt close
H	
+1	
R	
+1-	
11.63	
56	
0.0	
10-	

Figure B3. An example of a low quality crib sheet

It such fe	low y			
upper La	rels in Call Ans	4	ar is the	
Balan	ad			
		1 2 2 2		
	h + 1 children			
	when dise any			

Figure B4. Another example of a low quality crib sheet