Gender differences in the use of an online homework system in an introductory physics course

Gerd Kortemeyer*

*Lyman Briggs College and Division of Science and Mathematics Education, Michigan State University,
East Lansing, Michigan 48825, USA*
(Received 21 January 2009; published 26 May 2009)

The two genders make different use of being allowed multiple tries to solve online homework problems: male students frequently attempt to immediately solve the problem, while female students are more likely to first interact with peers and teaching assistants before entering answers. More male than female students state that they use the multiple allowed attempts to enter "random stuff," while more female than male students state that the multiple attempts allow them to explore their own problem solving approaches without worrying or being stressed out by grades.

I. INTRODUCTION

A gender gap in student performance in introductory physics courses is well documented and unfortunately persistent.[1] Closing this gap is an important goal of physics education research, but evidence as to the effectiveness of different teaching methods can be contradictory (for example, Lorenzo *et al.*[2] on the one hand and Pollock *et al.*[3] on the other hand came to different results regarding the effectiveness of interactive engagement methods on student performance on conceptual tests). It is therefore important to gain a deeper understanding as to the nature of the gender differences across various aspects of physics courses.[1] In this paper, we will focus on gender differences in the interaction with an online homework system.

In 1992, faculty members at Michigan State University started a computer-assisted personalized approach (CAPA) in order to provide randomized homework with immediate feedback for a small introductory physics course.[4–6] Different students are assigned different versions (different numbers, options, graphs, formulas, etc.) of the same problem, allowing them to discuss problems with each other without simply exchanging the solutions. Typically, students are allowed more than one attempt ("try") to solve formative assessment problems such as homework. Over the years, several studies regarding CAPA's educational effectiveness have been conducted (see Ref. 7 for a review), yielding a surprising yet unexplained outcome: female students appear to benefit significantly more from online homework than do male students, in terms of final course grade.

In this paper, we will not be able to prove any hypotheses regarding the differential benefit as that would require a comparative study of male and female students in otherwise similar courses with and without online homework. At Michigan State University, as well as most other user institutions, for several years all introductory courses have been taught with online homework. As online homework has been *proven* to be beneficial for learning success in large-enrollment introductory physics courses (i.e., courses large enough to deliver significant statistics), it would be hard to argue that solely for the purposes of an educational study, online homework should be eliminated for random subsets of students. Instead, all we can propose is to study gender dif-

ferences in the use of online homework in courses similar to those that have historically shown gender differences in performance and make inferences as to how these differences might affect learning success.

Section II describes previous findings regarding gender differences in using online homework, Sec. III describes the course under investigation, as well as the setup of the online system, Sec. IV describes the responses to an open-ended anonymous survey, Sec. V describes the outcomes of an analysis of online transaction data, and Sec. VI summarizes the results.

II. PREVIOUS RESULTS

In our previous studies, we consistently found online homework to be beneficial within our course settings and for our student population. In general, though, research regarding the effectiveness of online homework on student performance is inclusive and depends on many factors: course size, nature of the online homework, student population, level of the course, and teaching methods in other areas of the course. For example, Dufresne *et al.*[8] found a slight performance increase in some courses, while Bonham *et al.*[9,10] found no significant difference, and Pascarella[11] found counterproductive results.

A consistent pattern in our previous studies of large-enrollment introductory physics courses has been that online homework most effectively helps students who are on the brink of failing the course. A comparison between grade distributions before and after introduction of online homework typically shows that formerly strictly bell-shaped grade distributions get depleted around the 2.0 grade[6,7,12] (4.0 being the best grade, and any grade lower than 2.0 resulting in no credit for the course).

As an example, in an earlier study we examined a two-semester course, where in a particular year the first semester was taught without and the second semester with online homework.[12] Grades improved, but surprisingly, it turned out that this was mostly due to better performance of female students. In the first semester, the average grade of male students was 2.8 ± 0.8 versus 2.5 ± 1.1 for female students, i.e., more than a quarter grade difference. In the second semester, it was 2.8 ± 1.1 for the male and 2.8 ± 1.0 for the
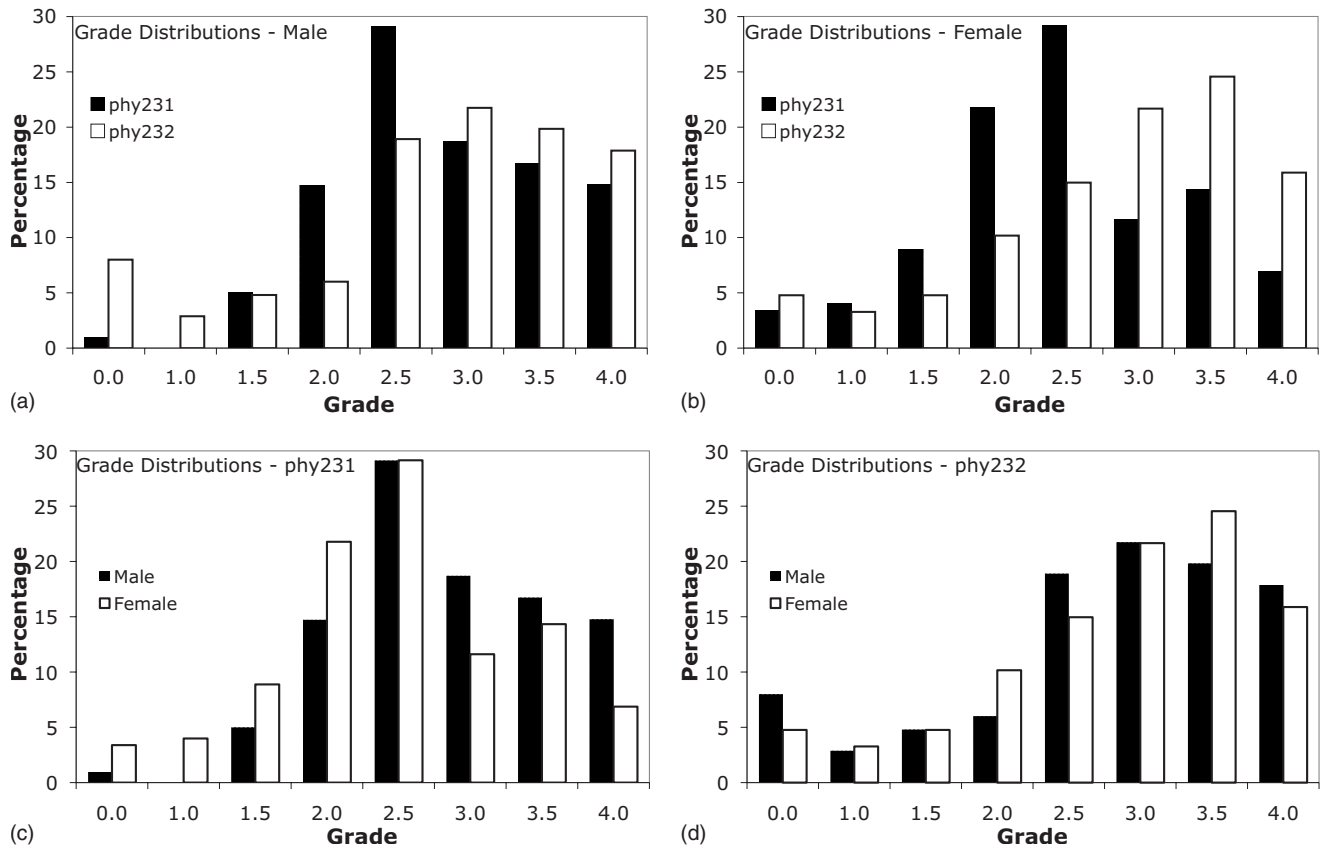
FIG. 1. Top row: grade distribution in the first semester (black, PHY 231) and the second semester (white, PHY 232) of a nonmajor introductory physics course. The left panel shows the grade distribution of male and the right panel of female students. Online homework was used in the second, but not the first semester (Ref. 12). Bottom row: grade distribution for male and female students within the same semester.

female students. It should be noted that online homework performance only minimally contributed to the course grade.

While standard deviations are given on these average grades to indicate their spread, they should not be interpreted in a sense of being significant or not: we are *not* measuring the average grades of one male and one female individual over and over; instead, these are hundreds of very different individuals. Also, the grading scale is designed to deliver a spread of grades between 0.0 and 4.0. It is thus more meaningful to compare the grade distributions.

Figure 1 shows the grade distributions for female and male students in the first and the second semesters. It can be observed that for female students, the grade distributions in the first semester (black, without online homework) and the second semester (white, with online homework) are very different, while for male students, there is hardly any difference between the semesters. Going back to the original data from Ref. 12, it was shown that the grade distribution of female students was significantly different from the male distribution in the first semester, indicating a significant gender gap, but this difference almost vanished in the second semester[7]—the female students had "caught up" to their male counterparts. A similar effect was observed at Central Michigan University.[13]

We were able to exclude simple attrition, population, or instructor effects [the enrollment fluctuations between the semesters were small compared to the overall enrollment, the grading was highly impersonal in this large-enrollment course, and the overall grade improvement was consistent with data polled across several years and sections of a majors course (for which, however, no gender information was available)[7]] but also unable to find alternative explanations for the difference in apparent effectiveness. In Ref. 13, only one gender difference in student interaction with online homework was discovered, namely, that females in semesters where they outperformed the male students usually did their online homework earlier, i.e., they were less likely to procrastinate until close to the due dates.

It should be acknowledged that the first and second semesters of the course deal with different subject matter, which suggests an alternative explanation for the closing of the gender gap: a significant portion of the first semester deals with mechanics—males might thus have an advantage due to typically greater involvement in sports, cars, etc. The second semester deals with the abstract topics of electricity, magnetism, and modern physics, where males would not have any such advantage. However, as the grades are not "curved," this explanation would suggest a closing of the gender gap by the male grades going down to the level of the female grades, which is not the case.
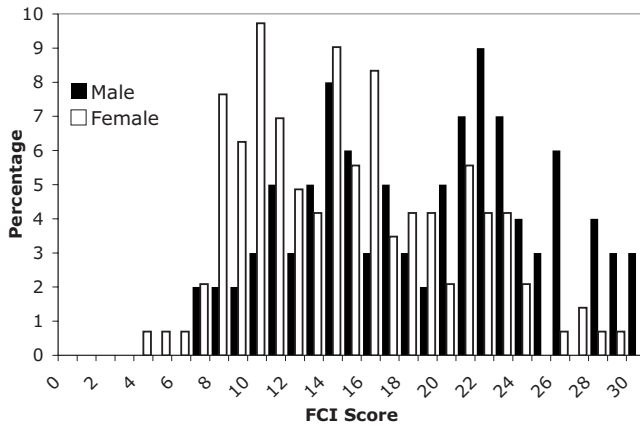
FIG. 2. Distribution of FCI scores for male and female students in the semester under consideration.

III. SETTING

The study took place in the first semester of an introductory physics course (mechanics) with mostly nonmajors and a high percentage of premedical students. For our study, we are using Maryland Physics Expectation Survey (MPEX) (Ref. 14) and Force Concept Inventory (FCI) (Ref. 15) data, exam scores, class grades, and results from a free-form survey on homework strategies, as well as transaction and performance data from the online homework system. The course had 97 male and 141 female students. Six tests (50% of the course grade) and a final exam (15% of the course grade) were administered over the course of the semester. Female students did consistently a few percent worse on tests than male students, but the differences (in the range from 1.3% to 9%) were statistically not significant. Similar to the setting described in Ref. 1, females did slightly better in participation "clicker" scores (5% of the course grade) and written recitation homework (15% of the course grade), and at the end of the semester, the average course grade of male and female students in the class was virtually identical.

Unfortunately, due to scheduling constraints, the MPEX (Ref. 14) and the FCI (Ref. 15) were not administered as pretest. Instead, those tests were administered after the third test in the seventh week of the semester, following instruction on force, momentum, and energy. The students received recitation attendance credit for participation. Male and female students both had an average MPEX "score" of 16 ± 6 (out of 34). The average FCI scores differed, though not statistically significant: male students had an average score of 18.8 ± 6 (out of 30), while female students had 14.8 ± 5. Figure 2 shows the FCI score distributions: there is a wide spread of scores, but the highest scores are almost exclusively achieved by male students, while the lowest scores are mostly from female students.

Homework problems, content pages, and other resources such as images, movies, and simulations (applets) were made available online, replacing a traditional textbook. Over the course of the semester, 422 online problems were assigned (15% of the course grade). 90% correctly solved homework would result in full homework credit, i.e., students get 100% homework credit for solving 380 problems, with no bonus for solving more. Numerical problems typically had 12 allowed tries, multiple-choice and other problems less, in correspondence to the number of answer options.

IV. SURVEY RESULTS

We administered an anonymous survey that asked students to report on several aspects of their use of the online homework. The students received recitation attendance credit for participation. The survey had 238 respondents, however, not all participants responded to all items. 138 participants identified themselves as female and 97 as male; 3 participants chose not to identify their gender (if everybody else reported their gender correctly, these would have had to be women, but no assumptions were made). 75% of the male and 74% of the female students stated that they had physics in high school, so there was basically no difference in high school physics preparation (high school preparation was found to be a predictor of the size of the gender gap in Ref. 1).

A. Time on task and "diligence"

An early hypothesis was that women are simply more "diligent" in their online homework and spending more time on task. At least the latter part of the argument can clearly be refuted: women reported spending 5.6 ± 3.3 h/week doing online homework, while men reported 5.1 ± 3.6 h/week—the half hour difference in the average is not significant. Also, analyzing the distributions of reported hours, no significant gender difference could be found.

B. Initial approach to problems

The open-ended sections of the survey yielded more significant results, in particular with regard to how students make use of the multiple attempts they had to solve problems. For this study, students were asked to write a short essay describing how they go about solving problems in the online system. Typical examples of student responses are as follows:

> "Try two or three times; read materials; try again; seek help from friends; attend help room."

> "First set up equation with intuition; if it doesn't work, refer to lecture or materials equations; third, see if hints help."

> "I try to use notes from lecture first. Then I go to the materials. If those don't work I sometimes use the internet, go to TAs, or look at the posted discussion."

> "(1) Read the corresponding materials, then try; (2) reread or recalculate; (3) ask friends for help; and (4) ask CAPA users for help."

The responses were then categorized by frequent actions and analyzed. Figure 3 illustrates the actions that the students reported as their initial approach to online problem solving.

The early hypothesis of a gender difference in diligence would suggest that instead of immediately attempting a prob-
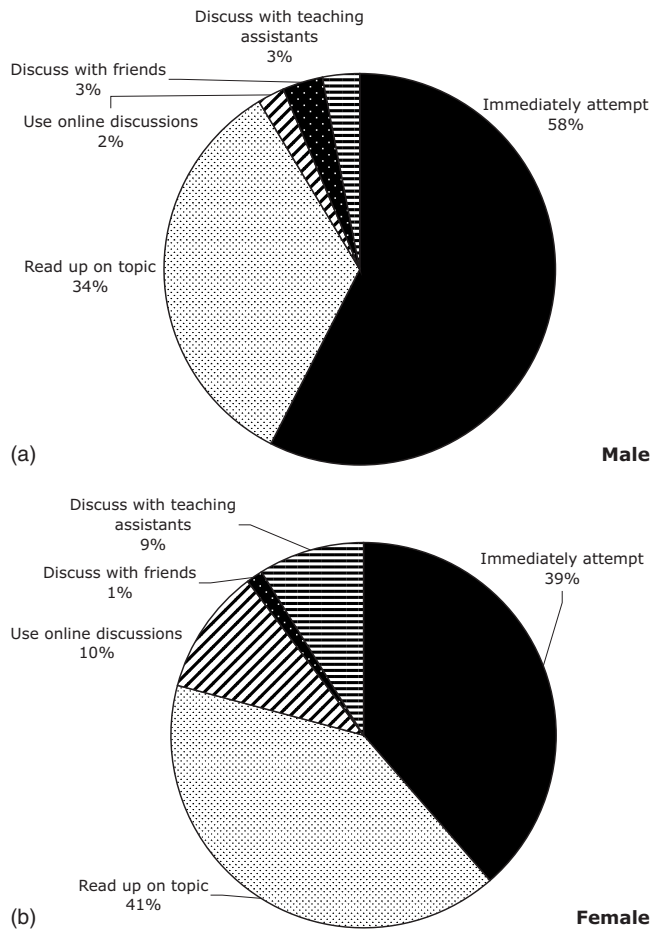
FIG. 3. First actions that students report taking on online homework problems. The top panel reflects the statements of male students, while the bottom panel reflects those of female students.

lem, women would first read up on the topic in the materials or in their class notes, however, the gender difference is not statistically significant: only 32 versus 62 male and 55 versus 80 female students first read up on the topic (34% male and 41% female; light gray sector in Fig. 3). The difference becomes even smaller when examining if reading the materials is at some stage part of the students' reported problem solving strategy: 69 versus 25 male (73%) and 105 versus 30 female (78%) students stated that they consult notes somewhere along the way of solving online physics problems. Thus, these simple measures of diligence can be rejected as significant behavioral differences.

On the other hand, the ratio of students reporting that they first turn to peers or teaching assistants before attempting the problems or reading up on the materials is higher for females (28 versus 107) than males (8 versus 86), 21% versus 9% ($p < 0.025$ for equal prevalence using Pearson's χ^2 test). Here, the strongest source of peer support appears to be the online discussions: 14 versus 121 female students first turn to the online discussion, while only 2 versus 92 male students report the same (10% female versus 2% male, $p < 0.05$). This result is consistent with a finding in an earlier year of the same course: female students on the average make 5.9 ± 1 online discussion postings over the course of a semester, while male students on the average only make 3.7 ± 0.7 con-

tributions [it should be noted that in the previous study we had no measure for viewing the discussions ("lurking")].[16] Thus, there is a highly significant difference in the initial approach to problem solving with regard to interaction with peers and teaching assistants, confirming the proverbial prejudice that "men do not ask for directions."

Among the male students, 54 versus 40 reported that the first thing they usually do is attempt the problem, while only 52 versus 83 female students did the same (black sector in Fig. 3), i.e., more than half of the male students (57%) immediately tried to solve the problems, while only 39% of the women did the same ($p < 0.01$)—men are significantly more likely to, before anything else, immediately attempt to solve the problem and only then resort to the materials or to asking peers and teaching assistants.

A future study might take the solution approaches extracted from the open-ended responses and turn them into a ranking format for detailed analysis and error estimation.

C. Use of multiple tries

Another open-ended section of the survey asked the question, "How do you make use of the multiple tries?" Some students made multiple statements, others none; examples of student answers are as follows:

"It helps when I make stupid mistakes like inputting a number wrong in my calculator."

"Trying random stuff; picking correct units; trial and error."

"I try to narrow my answer to two to three possible answers and guess all three."

"I like to know what I am doing wrong. So with multiple tries I can attempt to correct my mistakes."

"I only put in answers I believe are correct and check units. If tries get low, I get help ASAP."

"I like them because it allows us the freedom to struggle without stress attached. It also allows us to try more than one idea we may have."

Altogether, there were 111 statements by male and 181 statements by female students. These 292 statements were classified into 28 categories of similar statements. Due to the open-ended nature of the question, the categories are neither systematic nor necessarily mutually exclusive but also revealed unexpected results: among the more unproductive uses of multiple attempts is the category "I use tries for a process of elimination, often coordinated with friends." This apparently applies to randomizing multiple-choice, multiple true or false, and ranking questions, where students work in groups, with each one sacrificing a limited number of attempts (enough to still eventually "solve" the problem) to eliminate choices for the whole group. Another unproductive approach is "I vary the equations until they work," meaning, randomly changing signs, switching sine and cosine functions, multiplying and dividing by factors of 2, etc.

The percentage of prevalence of a certain category was calculated within each gender. For example, a statement to

TABLE I. Prevalence of statements regarding the question of how students make use of the multiple allowed tries to solve online homework.

Statement	Male (%)	Female (%)	Δ (%)
I submit random stuff or guess	14	8	−6
If I get it wrong, I backtrack my calculations	5	0	−5
Having many tries allows me to try different approaches and learn from my mistakes	22	18	−4
I use tries for a process of elimination, often coordinated with friends	5	3	−2
Having many tries does not allow me to just give up or quit	2	0	−2
Tries are useful to get answer formatting correct	3	2	−1
I don't use many tries	3	3	0
I take every try seriously	5	5	0
I vary the equations till they work	9	10	+1
Having many tries allows me to start working on the problem myself without waiting for help from others	2	3	+1
Tries are useful if I make errors in unit conversions	7	8	+1
Tries are useful if I make calculation errors	5	7	+2
Tries are useful if I make order of magnitude errors	1	3	+2
Tries are useful if I plug in the wrong numbers	0	2	+2
I use many tries	8	11	+3
Having many tries allows me to try out my own approach without the stress or worry about grades	6	10	+4
Other	3	7	+4

the effect of "I submit random stuff or guess" was made by 15 male and 14 female students, so its prevalence was calculated to be $15/111=14\%$ for male and $15/181=8\%$ for female students. The majority of the student statements converged into 16 categories that each had more than 1% prevalence for male or female students, i.e., were mentioned by more than two male or more than four female students. Table I shows these statements, sorted from statements more prevalent for male students down to statements more prevalent for female students.

Overall, there is very little difference between the prevalence of statements for male and female students. However, two differences are notable: 14% of the male but only 8% of the female students stated that they "submit random stuff," i.e., a tendency to waste tries in hope of getting the problem correct by chance. On the other hand, 10% of the female but only 6% of the male students stated that "having many tries

allows me to try out my own approach without the stress or worry about grades." Apparently, for female students more than for male students, having the possibility to correct mistakes makes them feel more comfortable and less anxious about using their own approaches.

A future study might take the more strongly distinguishing categories extracted from the open-ended responses and turn them into a Likert scale format for more detailed analysis and error estimation.

V. RESULTS OF AN ANALYSIS OF ONLINE TRANSACTION DATA

A. Access of online resources

The survey showed that female students are slightly (but not significantly) more likely to read up on the topic before attempting problems. An analysis of the actual online access patterns, however, shows that female students significantly more often access content pages versus problem pages than male students: male students accessed online content pages 49 854 times and homework problems 373 893 times (ratio 1:7.5), while for females, the access numbers were 76 585 versus 465 563 (ratio 1:6.1; $p<0.0001$). Of course we have no measure beyond self-reporting how often the students accessed external content resources, e.g., textbooks and lecture notes.

B. Resource access frequency

In previous sections, we focused on how often students access online content pages and online problems. It also matters how much time the students spend with each resource. It is difficult to realistically assess time on task since online access data only shows when new resources are being accessed, but there is no data on what happened in between: when there are 2 h between successive transactions, the student may have spent 2 h pondering the problem or the student may have stepped out to have a coffee. To get some handle on this, we looked at the time difference between two transactions of the same student on the same online resource, reflecting what may be called "access frequency." Examples for such transaction time differences would be the time between pulling up a problem and submitting an attempt, flipping to and from a resource, or the time between two subsequent attempts to solve a problem. The longer these times get, the more meaningless they are, as the student could have done an increasing amount of unrelated activities; on the other hand, for times of a few minutes, the student may indeed have stayed on task. There was a total of about 900 000 subsequent transactions on the same resource for male and 1 250 000 subsequent transactions on the same resource for female students.

Particularly times of less than 1 min indicate that the student did not invest much time into that particular resource or did not spend much time thinking about it between interactions. Figure 4 shows the distribution of times between subsequent transactions on the same resource for transactions that occurred within less than 1 min of each other. The first maximum of the male distribution and the maximum of the
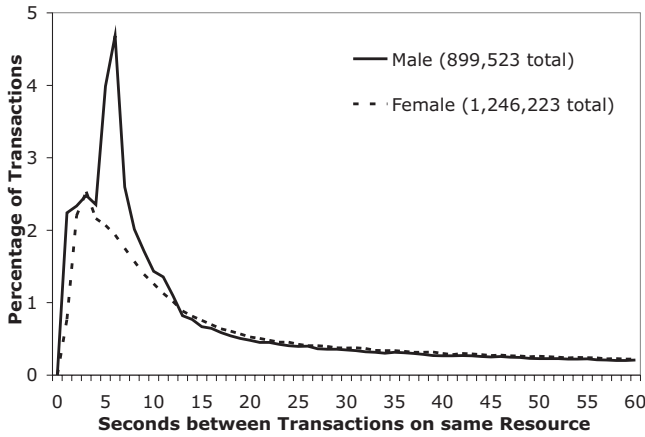
FIG. 4. Distribution of times between subsequent transactions on the same resource for transactions that occurred within less than 1 min of each other.

female distribution are both located at 3 s and both have the same value. These peaks probably represent just flipping past a page; the 3 s are likely due to server response time. The male distribution, though, has a second very strong maximum at about 6 s, likely corresponding to giving a page a cursory look for a few seconds.

Thus, looking at the total percentage of transactions that occurred in less than a given number of seconds (i.e., the integral of the distribution), the males are "ahead." Figure 5 shows the integrated number of transactions as percentage of the total. Of the total number of subsequent transactions on the same resource, for male students, about 45% occurred in less than 1 min, while for the female students, only about 38% did so. The females only slowly catch up, and 63% versus 57% of the transactions take less than 5 min. Male students were clearly spending less time with each resource.

C. Use of multiple tries

The survey data suggest that male students are more "wasteful" with the multiple allowed tries than female students, in that they are more likely to enter "random stuff." However, two immediate inferences are not true: that male
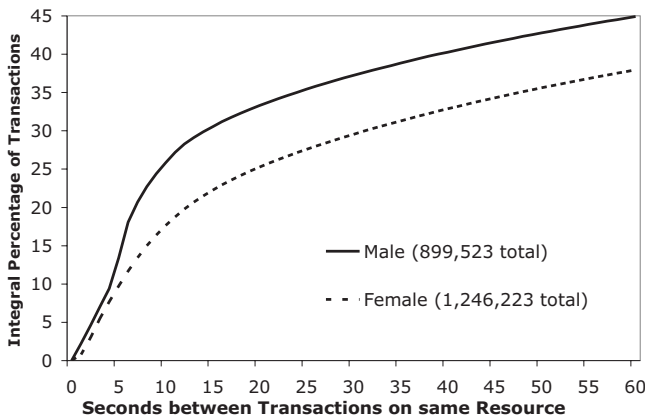

FIG. 5. Integral distribution of times between subsequent transactions on the same resource.
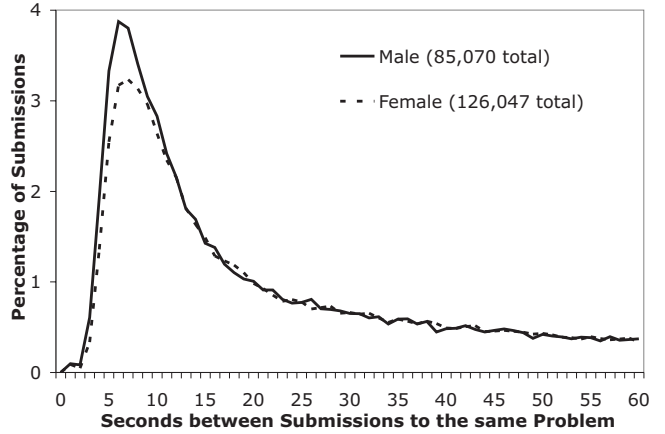
students would solve less problems and use more tries. Grading data show that male students on the average eventually solved 381 ± 40 problems, while female students solved 389 ± 30 problems. Both male and female students used an average of 2.1 ± 0.3 attempts for each solved problem. Also, the distributions are virtually indistinguishable: both male and female students on the average solved 220 problems in the first attempt and 80, 40, 20, and 10 problems in the second through fifth attempts, respectively.

D. Problem resubmission frequency

In this section, we investigate the time between repeated submissions to the same problem, i.e., after a failed attempt, how long do men and women take to submit another attempt? Figure 6 illustrates this. After a failed attempt, men most frequently submit another solution after 6 s, while women after 7 s; both times are rather short. Half of the resubmissions by male students occur within less than 38 s, while for female students, it is 45 s. 60% of all resubmissions by male students happen within 1 min, while for female students, the same percentage is 56%.

1. Single component models

Similar item response times are commonly modeled using a Weibull distribution,[17] as introduced in this context by Tatsuoka and Tatsuoka,[18]

$$
W_{t_0\lambda\beta}(t) = \begin{cases} \dfrac{\beta}{\lambda}\left(\dfrac{t-t_0}{\lambda}\right)^{\beta-1} \exp\left(-\left(\dfrac{t-t_0}{\lambda}\right)^{\beta}\right) & \text{for } t \geq t_0 \\ 0 & \text{for } t < t_0. \end{cases}
$$

$$(1)$$

Independent minimum-χ^2 fits ($0 \text{ s} \leq t \leq 60 \text{ s}$) yield $t_{0,m} = t_{0,f} = 4.01$ s for both male and female students, likely roughly corresponding to the server response time plus the minimal time to change an answer and resubmit. For male students, it is $\lambda_m = 59.60$ s and $\beta_m = 0.58$, and for female students, it is $\lambda_f = 72.72$ s and $\beta_f = 0.62$ ($\chi^2_m = 5.32$ and $\chi^2_f = 4.79$, respectively, $p > 0.995$). The Weibull distribution fails to satisfactorily model any gradual onset of the
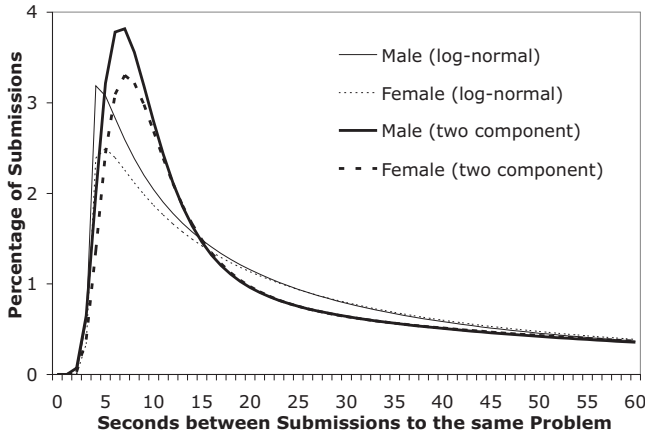

FIG. 6. Distribution of times between subsequent submissions to the same problem within less than 1 min of each other.

FIG. 7. Minimum-χ^2-fit single and two component logarithmic-normal fits of the data (Fig. 6).

TABLE II. Parameters obtained from independent minimum-χ^2 fits of Eq. (3) to the data (Fig. 6).

		Male	Female
	t_0	1.67	1.68
	p	25%	24%
Short	μ_S	1.9	2
	σ_S	0.58	0.58
Long	μ_L	4.21	4.33
	σ_L	1.44	1.39

distribution—while correctly modeling the long tail of the distribution, the onset of the distribution is completely determined by t_0 and thus modeled stepwise.

The data are better fit by a logarithmic-normal distribution,

$$L_{t_0\mu\sigma}(t)$$

$$= \begin{cases} \dfrac{1}{(t-t_0)\sigma\sqrt{2\pi}}\exp\left(-\dfrac{[\ln(t-t_0)-\mu]^2}{2\sigma^2}\right) & \text{for } t \geq t_0 \\ 0 & \text{for } t < t_0. \end{cases}$$

$$(2)$$

For male students, it is $t_{0,m}=2.96$ s, $\mu_m=3.62$, and $\sigma_m=1.85$; for female students, it is $t_{0,f}=2.95$ s, $\mu_f=3.79$, and $\sigma_f=1.79$ ($\chi^2_m=3.41$ and $\chi^2_f=3.37$, respectively, $p>0.995$). This model, though, under-represents the magnitude of the maxima and moves them to slightly too early times, as can be seen in Fig. 7.

Gamma and Poisson distributions completely failed to model the data ($\chi^2>50$).

The above distributions have been frequently and successfully used in literature to model response times in situations where subjects have only one chance at answering questions; due to initial hesitation, this may result in a much smoother onset of the distribution. In the scenario of this study, for both men and women, the fact that multiple attempts are granted might be the reason that there is much less hesitation to quickly submit another attempt.

2. Two component model

As suggested by the rapid onset of the distribution, the ability to submit multiple tries might lead to two different modes of answer submission: rapid submission of a random or guessed answer and a separate mode that follows traditional patterns. While Pascarella[11] warned that the availability of multiple tries might turn "thinkers" into "guessers," the same student might operate in both modes, depending on closeness of deadlines, number of remaining tries, or understanding of the subject matter. Both modes, the long ("thinking") and the short ("guessing") mode, would operate on

different time scales but would have superpositioned distributions. To model this behavior, a two component logarithmic-normal distribution [Eq. (2)] was investigated,

$$T_{p,t_0\mu_S\sigma_S\mu_L\sigma_L}(t) = pL_{t_0\mu_S\sigma_S}(t) + (1-p)L_{t_0\mu_L\sigma_L}(t). \quad (3)$$

Here, the parameter p determines the prominence of the short (guessing) mode. Independent minimum-χ^2 fits were performed for both male and the female distribution data, Table II shows the obtained values, and Fig. 7 illustrates the outcome. With such a large number of free parameters, it is not surprising that excellent fits ($\chi^2<0.2$) could be obtained, but it is surprising how similar the parameters for the short components are: in this model, both men and women appear to engage in the same amount of guessing behavior (about 1/4 of the submissions) on the same time scale; the distributions differ in the long (thinking) component. Within this two component model, men and women are equally likely to submit random stuff (thus not confirming the self-reported behavior in Sec. IV C), but, when actually more seriously considering the problem, women invest a little more time, and their submission frequency distribution is more widely spread.

VI. CONCLUSIONS

Prompted by the gender-differential effectiveness of online homework in large-enrollment introductory physics courses, we examined how male and female students interact differently with an online homework system in a similar setting. The course under investigation in this study, however, only exhibited a small gender-gap in test and FCI performance to begin with.

Overall, remarkably, similarities and null results far outweigh the differences in both self-reported and analyzed characteristics of the use of the online homework system. We did not find any one strongly distinguishing feature, however, we found several small differences, the sum of which might result in differential benefits.

Male and female students make different use of being allowed multiple tries to solve online homework problems: male students frequently, before anything else, attempt to immediately solve the problem, while female students are more likely to first interact with peers and teaching assistants before entering answers. More male than female students state that they use the multiple allowed attempts to enter

random stuff, while more female than male students state the multiple attempts allow them to explore their own problem solving approaches without worrying or being stressed out by grades. Male students are slightly more quickly clicking around when working through online materials and are slightly more quickly resubmitting new attempts after previously failing to solve a problem. A two component analysis showed that the problem submission frequency may in fact be the superposition of two distinct modes: a short turnaround mode of likely submitting random stuff or guessing, which male and female students are equally likely to engage in, and a longer turnaround time associated with a deeper consideration of the problem, for which women are investing slightly more time than men.

ACKNOWLEDGMENTS

*URL: http://www.lite.msu.edu/kortemeyer/

[1] L. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. **5**, 010101 (2009).

[2] M. Lorenzo, C. H. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, Am. J. Phys. **74**, 118 (2006).

[3] S. J. Pollock, N. D. Finkelstein, and L. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?, Phys. Rev. ST Phys. Educ. Res. **3**, 010107 (2007).

[4] E. Kashy, B. M. Sherrill, D. T. Y. Tsai, D. Weinshank, M. Engelmann, and D. J. Morrissey, CAPA: An integrated computer assisted personalized assignment system, Am. J. Phys. **61**, 1124 (1993).

[5] E. Kashy, S. J. Gaff, N. Pawley, W. L. Stretch, S. Wolfe, D. J. Morrissey, and Y. Tsai, Conceptual questions in computer-assisted assignments, Am. J. Phys. **63**, 1000 (1995).

[6] M. Thoennessen, E. Kashy, Y. Tsai, and N. E. Davis, in *Proceedings of the 31st Hawaii International Conference on System Sciences* (IEEE, New York, 1997), Vol. 1, pp. 231–239.

[7] G. Kortemeyer, E. Kashy, W. Benenson, and W. Bauer, Experiences using the open-source learning content management and assessment system LON-CAPA in introductory physics courses, Am. J. Phys. **76**, 438 (2008).

[8] R. J. Dufresne, D. Hart, J. P. Mestre, and K. Rath, The effect of web-based homework on test performance in large enrollment introductory physics courses, J. Comput. Math. Sci. Teach. **21**, 229 (2002).

[9] S. Bonham, R. Beichner, and D. Deardorff, Online homework: Does it make a difference?, Phys. Teach. **39**, 293 (2001).

[10] S. Bonham, D. Deardorff, and R. Beichner, A comparison of student performance using web- and paper-based homework in college-level physics, J. Res. Sci. Teach. **40**, 1050 (2003).

[11] A. Pascarella, in Proceedings of the NARST 2004 Annual Meeting, Vancouver, BC, 2004 (unpublished) (http://www.loncapa.org/papers/204416ProceedingsPaper.pdf).

[12] D. A. Kashy, G. Albertelli, E. Kashy, and M. Thoennessen, Teaching with aln technology: Benefits and costs, J. Eng. Educ. **89**, 499 (2001).

[13] P. Kotas, Master thesis, Central Michigan University, 2000.

[14] E. F. Redish, R. N. Steinberg, and J. M. Saul, Student expectations in introductory physics, Am. J. Phys. **66**, 212 (1998).

[15] I. Halloun, R. R. Hake, E. P. Mosca, and D. Hestenes, URL: http://modeling.la.asu.edu/R&E/Research.html

[16] G. Kortemeyer, An analysis of asynchronous online homework discussions in introductory physics courses, Am. J. Phys. **74**, 526 (2006).

[17] W. Weibull, A statistical distribution function of wide applicability, ASME Trans. J. Appl. Mech. **18**, 293 (1951).

[18] K. K. Tatsuoka and M. M. Tatsuoka, *Technical Report* (University of Illinois at Urbana, Champaign, 1978).