Assessing Thinking Skills in Free-response Exam Problems: Pandemic Online and In-person
Fatema Al-Salmani, Beth Thacker and Jordan Johnson, Texas Tech University

Abstract
We present an analysis of students’ thinking skills as evidenced in free-response exam problems. We compare two sections, one taught online and the other in-person during the pandemic. The rubric was originally designed based on Bloom’s taxonomy (revised version) to compare thinking skills of students in classes taught by different pedagogies. We discuss the instrument, present results and interpretations.

Introduction
In recent years there has been a move toward recognition of the need to assess not just conceptual understanding, but also students’ thinking skills. There has been significant work done by a number of groups both in the development of assessment instruments and on promoting thinking skills in the classroom [1-8]. Most of these authors define and discuss “critical thinking” in a particular context. We are interested in assessing the thinking skills employed by students in introductory physics courses as evidenced in free-response (FR) exam questions. A rubric based on Bloom’s taxonomy (revised version) [9] has been designed for this purpose [10]. Here we are focus on comparing the thinking skills of students taught online versus students taught in-person during the pandemic in algebra-based physics classes.

Methods
We chose some questions from the exam problems that had been administered across the online physics-I course and applied them through the in-person course. Four common exam questions were used. Two example questions are presented in this poster.

The Rubric
Bloom’s taxonomy levels for physics-free-response exam problem solutions:
0) None
i) Nothing
ii) Totally incorrect
iii) Statement only
iv) Partial Remember – includes some facts or concepts needed to solve a problem, but at least one necessary concept or fact is missing or incorrectly remembered.
1) Remember: Recall facts and basic concepts
Solution includes recollection of a fact or concept or set of facts or concepts needed to solve the problem in either mathematical or written form. (If one or more necessary concepts are missing, it counts as Partial Remember and not Remember.)
2) Understand: Explain ideas or concepts
Solution includes a correct application of the facts or concepts recalled. This could be verbal or mathematical, it includes any of the definitions, constraints given in the particular problem and how they would be applied. A picture must be supported by a verbal explanation.
3) Apply/Analyze: Use information in new situations/draw connections among ideas
Solution demonstrates a valid, logical and consistent process for determining an answer, applying (1) and (2), and demonstrates evidence of understanding an underlying model or fundamental principle.
4) Evaluate: Justify a stand or decision
Solution includes a comparison of or discrimination between different possibilities or an explanation of why an answer is correct or makes sense or why it is incorrect.

Example Problem 1
2A boy is dragging a 15kg box across the floor with a rope at an angle of 54° above the horizontal. He pulls with a force of 45.0 N. The box accelerates at 0.5m/s².
1) a) Is the normal force equal to the gravitational force? Show work and/or explain your reasoning
b) Determine the coefficient of kinetic friction between the box and the floor. Show your work.
c) Is there a coffee mug on top of the box, would it be possible for the mug to remain on top of the box, or would the box slide out from under it? Show your work and explain your reasoning.

Results
In Figs. 1 and 2, we show the results of the comparison of students taught in two different settings. Fig. 1 shows the results of Problem 1 and Fig. 2 shows the results of Problem 2. In Problem 1 parts (a), (b) and (b), the students in the online course demonstrated higher level thinking skills more often. In the other parts there were no significant differences between the courses based on Fisher’s exact test. The results of Fisher’s exact test for Problem 1 were p = 0.0433, p = 0.008 and p = 0.5797 for parts (a), (b) and (c) respectively. This indicates that the results (a) and (b) are significantly different, but not for part (c). For Problem 2 the results were p = 0.0048, p = 0.5489, p = 0.5276, p = 0.00 and p = 0.0005 for parts (a), (b), (c), (d) and (e). Based on this, the results of parts (a), (d) and (e) are significantly different.

For interrater reliability, we used Cohen’s kappa to compare any two raters. Because we observed that the rubric levels were not hierarchical (if a level is coded with a 1, all of the levels below it are also coded 1), the weighted Cohen’s kappa is the more relevant statistic than the unweighted. We report both the weighted and unweighted results for each problem. The weighted Cohen’s kappa was always greater than 0.80 and the unweighted was always greater than 0.70 when any two of the problems were compared for Problem 1. For Problem 2, the weighted Cohen’s kappa was always greater than 0.80 and the unweighted was always greater than 0.60 and usually greater than 0.70.

Conclusion
We have presented the results from two sample problems, as evidenced by our rubric, of the thinking skills of students in two pandemic courses. One course was taught online and the other in-person. According to the results of the problems we ran, there was no evidence of significant differences in thinking skills between students taught online and in-person. This is in contradiction to our preception.

Despite our perception that the online pandemic course was frustrating and uncomfortable for students and faculty, it is possible that we overlooked the small groups and student/instructor interactions similar to the in-person course, was responsible for the lack of significant change in thinking skills.

In conclusion, no significant differences were found in thinking skills between students taught in the online and in-person pandemic courses.

References