International Journal of Engineering Education Vol. 37, No. 5, pp. 1414–1428, 2021
Printed in Great Britain
0949-149X/91 $3.00+0.00

Comparing Engineering Problem-Solving Ability and Problem Difficulty Between Textbook and Student-Written YouTube Problems*

UCHENNA ASOGWA[1], T. RYAN DUCKETT[2], AMANDA P. MALEFYT[3], LINDSEY STEVENS[1], GALE MENTZER[2] and MATTHEW W. LIBERATORE[1]
[1] Department of Chemical Engineering, University of Toledo, Toledo, OH, USA.
[2] Acumen Research and Evaluation, Toledo, OH, USA.
[3] Dr. John J. McKetta Dept. of Chemical & Bioprocess Engineering, Trine University, Angola, IN, USA.
E-mail: uchenna.asogwa@rockets.utoledo.edu, timothy.duckett@rockets.utoledo.edu, malefyta@trine.edu, lindseystevens2016@gmail.com, gmentzer@acumenresearcheval.com, matthew.liberatore@utoledo.edu

Problem solving is a signature skill of engineers. Incorporating videos in engineering education has potential to stimulate multi-senses and further open new ways of learning and thinking. Here, problem solving was examined on problems written by previous students that applied course concepts by reverse engineering the actions in videos. Since the videos usually come from YouTube, the student-written problems are designated YouTube problems. This research focused on examining the rigor of YouTube problems as well as students' problem-solving skills when solving YouTube problems compared to Textbook problems. A quasi-experimental, treatment/control group design was employed, and data collected was evaluated using multiple instruments. NASA Task Load Index survey was used to collect ~1200 ratings that assessed rigor of homework problems. Problem-solving ability was assessed using a previously-developed rubric with over 2600 student solutions scored. In the treatment group where students were assigned ten Textbook and nine YouTube problems, students reported an overall similarity in rigor for both YouTube and Textbook problems. Students in the treatment group displayed ~6% better problem solving when completing YouTube problems compared to Textbook problems. Although higher perceptions of problem difficulty correlated with lower problem-solving ability across both groups and problem types, students in the treatment group exhibited smaller decreases in problem-solving ability as a result of increasing difficulty in the Textbook problems. Overall, student-written problems inspired by YouTube videos can easily be adapted as homework practice and possess potential benefits in enhancing students' learning experience.

Keywords: problem solving; visual communication; problem-based learning; student perception

1. Introduction

As of June 2018, over four billion people had access to the Internet, which represented about 55% of the world's population [1]. Almost all current undergraduate students began interacting with digital technology at a young age and today many everyday tasks revolve around the utilization of electronic devices such as cell phones, tablets, and computers. These students are often referred to as digital natives [2]. Nearly instant access to course-related information, such as looking up unit conversions, finding physical properties, or verifying an equation, offer technology-savvy students some advantages in learning course content. Some learning style differences are being identified between digital natives and past generations. In many cases, digital natives show a preference for visual media compared to text, are strongly motivated by projects having a real-world component, and possess shorter attention spans [3].

Homework problems from textbooks allow students, especially in engineering, to practice problem solving. However, solutions manuals are often available on the Internet, so students can locate and copy the correct solution while putting little effort into learning new material or developing problem-solving skills [4, 5]. Copying solution manuals as a form of studying can inhibit success in a course [5]. Therefore, finding new ways to develop interesting and textbook-quality homework problems to both engage and educate digital native students is a central theme of this work.

Recent surveys predicted that between 2015 to 2020 more than 36% of jobs across all industries require complex problem-solving as a core skill [6]. Not only is complex problem solving relevant in today's workspace, complex problem-solving skills are predicted to be the most prevalent skill to thrive in the workforce in 2030 [7]. Most instructional approaches limit students' ability to transfer learning by focusing on only course-specific information. Accrediting Board for Engineering and Technology (ABET) standards emphasize problem solving and knowledge of current issues; Infusing real world situations into engineering education helps students' understanding become more integrated [8, 9]. Therefore, tying engineering problem solving

* Accepted 30 May 2021.

with real world environments aligns well with current and future workforce needs.

In addition to real world situations, senses play a vital role in learning. Vision generally creates both short term and long-term memories more effectively than the other four senses [10]. Visual representation is an important part of successfully solving complex problems. Visual learning methods open new ways of problem solving and thinking, as well as enhance the education and practice of science and engineering [11–17]. In addition, the seemingly endless information on the Internet, and specifically YouTube videos, provide an array of contexts to connect engineering fundamentals to visual situations, which can be motivating and interesting. Therefore, the engagement and productive learning from searching for, identifying, watching, and translating YouTube videos ties in well with cutting-edge research in neuroscience and learning science [10, 18, 19].

Active learning and student-centered pedagogies lead to improved learning compared to traditional teacher-centric techniques, such as lecture [20, 21]. Also, involving students' enthusiasm is advantageous to learning [22]. Pedagogies are adapting to current students' strengths by integrating their digital habits into the higher-education classroom. In fact, technology in the classroom is expected by many digital natives (e.g., clickers, tablets, just-in-time teaching, YouTube) [5, 14, 23–27]. Implementation of technology as a form of active learning is a useful approach that connects students and learning [28, 29]. Therefore, engaging the current generation of students using visual technology mediums, like YouTube, in a positive way was one motivation directing this project.

Originally called YouTube Fridays, the YouTube pedagogy, which is explored here, started as a way to introduce and engage students in thermodynamics and material and energy balance courses. The first five minutes of Friday's classes were dedicated to course-related videos selected by students. As a result the vast majority (> 80%) of students affirmed a better understanding of the field of chemical engineering [27]. In subsequent semesters, students selected YouTube videos and created engineering problems related to the course material. Positive feedback in areas related to real-world connection and problem solving confidence were recorded [23]. Videos continued to be taken from YouTube or other websites in the public domain. Hundreds of student-written problems (hereafter referred to as YouTube problems), have been created in recent years [14, 23, 30]. While the writing is largely open-ended, a small number of boundaries keep the students' authoring focused. The assignment is initiated by students selecting a YouTube video to reverse engineer. From the video, students write a course-related problem to be complete, correct, and appropriately difficult to assign as a homework problem for the course.

YouTube pedagogy which deploys a strategy where students apply course concepts to reverse engineer YouTube videos and create new homework-quality problem statements and solutions is built upon sound learning theories about engaging and motivating students through constructive learning activities. Moreover, constructive activities can promote cognitive processes related to problem solving skills [4, 31, 32]. These skills include new ways of conceptualizing and organizing information, integration of new information with existing knowledge, and repairing misconceptions which can also apply to real world problems [31]. Thus, assessing problem-solving ability on Textbook and YouTube problems expands upon previous work. One strategy to measure problem solving is through a performance rubric which provides instructors with valid and reliable information to monitor and offer feedback on students' progress related to specific criteria [33, 34]. An example of a performance rubric is PROCESS (Problem definition, Representing the problem, Organizing information, Calculations, Evaluation, Solution presentation and Self-assessment). PROCESS was developed to measure the conceptual and analytical skills required when problem solving in an engineering class [35–39]. PROCESS was designed to track each step involved in solving problems in real time collected using tablets. Since PROCESS had been used in engineering courses and on problems based on real-world scenarios similar to YouTube problems, the PROCESS rubric was adapted for the current study. The rubric will be discussed further in the methods section.

One common practice for assessing problem difficulty is by making judgments based on an instructor's experience which is limited in ability to provide a quantifiable measure of problem difficulty [40]. Assessing problem difficulty through item analysis which estimates the probability of successful problem solutions based on student performance, provides a quantifiable measure [41]. However, assessing problem difficulty based solely on performance measures does not consider the presence of extraneous factors that could have influenced success rates [40]. Another measure of problem difficulty is based on students' perception. Three widely used self-reported measures of mental workload are the Modified Cooper-Harper Scale, NASA-Task Load Index (NASA-TLX), and Subjective Workload Assessment Technique (SWAT) [42, 43]. The current study adopted NASA-TLX because of its ease in administration, which is detailed in the methods section.

Therefore, this study examines the problem solving and problem difficulty in a new context, i.e., when comparing student-generated YouTube problems and Textbook problems in an undergraduate engineering course. Research questions seek to examine the rigor and efficacy of YouTube problems by investigating the effects that solving YouTube problems have on students' problem-solving ability. Results may inform educators as to an engaging means of providing students with problem-solving practice.

2. Methodology

Student-written YouTube problems fall under a category of contextual or authentic problems that possess the potential of improving learning outcomes [44]. Research questions revolve around probing the influence of solving YouTube problems with respect to problem solving ability as well as students' perception of problem difficulty. The methods section begins by discussing the features of YouTube problems, relevant course topics covered by the problems, and a description of the participants. Further subsections cover the deployment of various tools in collecting data pertaining to problem solving and problem difficulty. Finally, the statistical approaches are summarized.

2.1 YouTube Problems

YouTube problems are student-written, homework-style problems formed by reverse engineering a video to apply course concepts. YouTube problems possess features that examine student's learning at numerous levels of Bloom's taxonomy [45]. Examples of YouTube problems are detailed in a number of publications [14, 23, 27]. YouTube problems can be implemented in class, as part of homework sets, or in quizzes/exams, but this study limited deployment of YouTube problems to homework. Specifically, YouTube problems are close-ended problems with quantitative answers analogous to Textbook problems. By incorporating values from a video, the theme and scope of YouTube problems varies greatly, from mimicking Textbook problems, problems with single questions, problems with multiple parts, and sets of conceptual questions.

A How It's Made video for Nylon production inspired a problem for a reacting system. The problem statement is similar in length to an average Textbook problem (Fig. 1) and includes a balanced chemical reaction, multiple parts/questions, and a process flow diagram. The idealized reaction and

Watch the video titled: Nylon Production
https://www.youtube.com/watch?v=4GxeSO7DyaE

The reaction between two monomers, adipic acid ($C_6H_{10}O_4$ = "A") and hexamethylenediamine ($C_6H_{16}N_2$= "H"), produces nylon ($C_{12}H_{26}N_2O_4$ = "N"). The total flow rate of the reactor feed stream is 385 mol/hr and contains an equimolar mixture of adipic acid and hexamethylenediamine. The reactor effluent contains A, H, and N. The equilibrium constant is 43.4.

a) Draw and label a process flow diagram.

b) Calculate the component molar flow rates (mol/hr) of the reactor effluent stream.

c) If the equilibrium constant is increased by 17.2%, will the flow rate of adipic acid exiting the reactor increase, decrease, or stay the same?

(a)

Exercise 3.3.2: Methanol reactor. About

The synthesis of methanol from carbon monoxide and hydrogen includes nitrogen as an inert carrier gas. The feed to the reactor is 425 mol/min with 102 mol/min CO, 0.143 mol fraction of N_2, and the balance H_2. In the reactor, a single-pass conversion of CO is 75.8%. The reactor effluent goes to a condenser for further separation.

(a) Draw and label a process flow diagram and number the streams.

Solution

(b) Calculate the component molar flow rates for all of the components exiting the reactor (mol/min).

Solution

(c) The conversion of CO increases by 8%. Will the mole fraction of nitrogen exiting the reactor increase, decrease or stay the same?

Solution

(b)

Fig. 1. Student written reaction problem (a) and a typical Textbook problem statement (b) for material and energy balances assigned to students as homework.

separation scheme are common in this course. The first part of the problem statement for 'Nylon production' (Fig. 1) is an example of the interpretation type problem, which requires that a process flow diagram be drawn, and all streams labeled. The attention to detail in identifying process units and streams from the video is part of the problem-solving process used throughout the course.

The YouTube pedagogy was implemented in a material and energy balances course, which is an introductory freshman/sophomore course in most chemical engineering programs. We employed a quasi-experimental, treatment/control group design (Table 1). Random selection and assignment within a single group of students were not considered due to only one section of the course being offered per year. However, a similar population of students at another university who were studying the same course content and using the same textbook, but not employing YouTube problems, served as a control group. Faculty teaching the respective groups had previous experience teaching the course and they collaborated to ensure similar content delivery and used the same control (i.e., Textbook [46]) problems. Homework problems assigned to students covered a range of course topics (Table 2); see Table A.1 for detailed information on each problem.

While many problems that were not part of the current study were completed by students, we considered two possible conditions – Textbook homework (traditional homework problems) and YouTube problems. YouTube problems were written by previous students and assigned to current students as homework problems. Instructors selected the YouTube problems by mapping concepts to the course syllabi. Before administration of YouTube problems, instructors proofread the problems and sometimes reworded problems to ensure that the language in the problems was clear. YouTube problems were implemented for three course topics, namely material balances with reactions, material balances with reaction and recycle, and material balances for multiphase systems (Table 2). While video links were included with all of the YouTube problems, solving YouTube problems is possible without watching the video. Video views for the treatment group were not documented in this study.

For the treatment group, homework assignments, nominally 3 to 5 problems per week, varied between only Textbook problems, only YouTube problems, and a combination of Textbook and YouTube problems. During the initial weeks of the study, both groups solved only Textbook problems as a measure of group equivalency. Students' hand-written solutions were scanned solutions and

Table 1. Summary of problem assignment to treatment and control group

Group	YouTube	Textbook	Class size
Treatment	9	10	90
Control	0	10	23

Table 2. Number and type of homework problems assessed in topics of a material and energy balances course

Topic	Textbook	YouTube
Mass and mole fraction calculations	1	0
Non-reacting material balances	2	0
Material balances with reactions	1	3
Material balances with reaction and recycle	1	3
Material balances for multiphase systems	1	3
Non-reacting material balance and STP	1	0
Energy balance	2	0
Transient material balance	1	0

Table 3. Highest completed mathematics course by group

Math Course	Control (%)	Treatment (%)
Calculus 1	8	68
Calculus 2	52	12
Calculus 3	30	12
Differential Equations	9	3
> Differential Equations	0	4

scored anonymously using PROCESS after the course's completion.

The intervention constituted of a treatment group of 90 students (41% female) from a large public university and 23 students (22% female) in the control group at a private university. The control group consisted of second-year students who learned the course material over a two-semester period unlike the treatment group that occurred in the students' first year of study and covered material and energy balances course over one semester. In order to balance sample sizes and reduce problem scoring burdens, we randomly selected ~30 students' work from the treatment group to be scored using the PROCESS instrument. The different distributions for highest mathematics courses completed by group (Table 3). This difference can be explained by the course sequence noted above.

2.2 Assessing Problem-Solving Ability Using PROCESS

Students' problem-solving skills were measured using a modified PROCESS rubric with 6-stages: Problem definition, representing the problem, organizing information, calculations, solution completion and accuracy (see Table A.2 for PROCESS rubric). PROCESS evaluates both the problem-solving process and the final solution(s) (see Fig. A.3 for example of detailed scoring). PROCESS was modified to assess the problem-solving process for solved handwritten homework problems, which differs from its original use where participant solutions were collected on Tablets and custom software could detect erasing and other details [37, 39]. The tool was modified to suit material and energy balance problems [47]. Each item in the revised PROCESS consists of four scaling levels ranging from 0 to 3 with zero being the minimum attainable score.

Prior to scoring with the modified PROCESS, identifiers regarding student or group identity were removed. Participants' names were replaced with a project-assigned ID number to maintain privacy and to mask group membership, i.e., treatment or control group, from raters. All students' solutions were scored using the PROCESS rubric after the semester. Thus, PROCESS scores did not reflect or have an effect on students' course grades. Also, correct solutions, and similarly PROCESS scores, for YouTube problems did not require watching the linked video.

In the present analysis, four different raters used the PROCESS tool to assess problem solving. Raters' assessments were analyzed to determine how consistently raters measured problem-solving ability. Traditional statistical (intraclass correlation coefficient, ICC) and item response measures (rater severity from the Rasch many facets model) of inter-rater reliability were computed for the four raters, as previously described [48]. The many-facet Rasch measurement model provided a correction for any differences in rater severity in assessing PROCESS scores, such that the scores were free from any rater bias/leniency (see Fig. A.4 for initial inter-rater assessment) [49]. A previous paper detailed the process of establishing inter-rater reliability for multiple raters using the PROCESS rubric[48]. Consequently, the intraclass correlation coefficient (ICC) reported that the scores from the four raters were highly reliable. The average measure ICC was 0.92 with a 95% confidence interval from 0.90 to 0.93 (F (262, 786) = 11.8, p < 0.001).

Given the discrepancy between the control and treatment groups with regard to highest level of math courses, a Spearman's rank-order correlation was conducted between PROCESS score per student and level of math. A weak positive correlation between Textbook PROCESS score and level of math was found (rs = 0.26, p = 0.011). Therefore, in order to control for the significant difference math level had on PROCESS score, differences in group problem solving between the treatment and control groups were tested using ANCOVA. The level of statistical significance was set *a priori* at $p \leq 0.05$.

2.3 Assessing Problem Difficulty with the NASA TLX

In the case of problem solving, researchers must know how difficult the problem is in order to make a valid assessment of performance, i.e., comparing performance across problems, problem types, and participants. NASA TLX (Task Load Index) provides an appropriate gauge of problem difficulty [40]. For over three decades, NASA TLX has measured workload by assessing six constructs: three measuring demand put on the participant by the task, and three measuring stress added by the participant as a result of interacting with the task. The three measures of task demand are mental demand, physical demand, and temporal demand while stress measures include effort, performance, and frustration (see Table A.5 for a list of NASA TLX questions). The original NASA TLX measured workload in two stages consisting of participants ratings of each subscale and a pairwise comparison of each subscale [40, 50-53]. For ease of administration, NASA TLX could utilize participants' rating in exclusion of the pairwise comparison of subscales, which is often referred to as Raw TLX (RTLX) [54].

The current study utilized only the participants' TLX rating to measure the rigor of problems (Table 4). NASA TLX was modified such that the original 21-point sliding scale was reduced to a 6-point rating scale, where 1 is the least difficult and 6 the most demanding This change reduced the number of response options to increase the precision of the students' ratings since previous literature has found that including seven categories or more frequently exceeds the discriminative capacity of the respondent [55]. For each participant, responses to the 6 TLX questions were analyzed using the Rasch measurement model (discussed below) and rescaled to an aggregate rating of overall problem rigor that ranges from 0 to 100. More demanding tasks earn higher scores. Difficulty of a problem was assessed by averaging participants TLX scores for each problem. Analysis compared overall problem difficulty for different problem types and consistency in group responses.

2.4 Rasch Measurement Model

Rasch analyses of the PROCESS and TLX data were conducted using the Rating Scale model [56] in WINSTEPS (version 4.5.2, Beaverton, OR) [57]. This approach converted the ordinal-level, raw scores from the instruments into interval linear measures required for other statistical analyses. In brief, an iterative version of the PROX method provides starting values for the joint maximal likelihood estimation of the free parameters (person ability, item difficulty, and k-1 threshold calibrations). This procedure builds off a stochastic Guttman pattern that posits as items increase in difficulty, they require higher ability on the part of the student in order to succeed on the item. The ability of the parameters estimated in the Rasch analysis (ability of students and difficulty of items) to explain variance in the observed scores provides evidence for construct validity, i.e., the extent to which we are measuring "problem-solving ability" and not a different construct [58]. Thus, raw scores are transformed into an information weighted probabilistic parameter: the joint logarithmic calculations of parameters based on the convergence of observed scores onto those expected by the model entail independence from specific person and item distributions [59].

2.5 Relationship Between Problem-Solving Ability and Problem Difficulty

Several linear regression models were tested using IBM SPSS Statistics (version 24, Armonk, NY) to examine the extent student perceptions of problem difficulty predicted their problem-solving ability. The predictor variable – perceptions of problem difficulty – came from the NASA TLX. The response variable – problem-solving ability – came from the PROCESS scores. Correlation between the observed and predicted values of the predictor variable as well as proportion of the variance in the predictor variable that could be predicted from the response variables were consulted in addition to overall model significance [60].

3. Results and Discussion

YouTube pedagogy is a constructive learning activity involving visuals, which can enhance short- and long-term memory formation [4, 15, 31]. Considering YouTube problems as an alternative to Textbook problems, the central hypothesis is that student-generated YouTube problems promote better problem-solving skills than traditional Textbook problems. Thus, evaluating the efficacy of YouTube problems addressed two primary research questions:

1. Does solving YouTube problems improve students' problem-solving skills compared with solving problems from textbooks?
2. Are YouTube problems and Textbook problems perceived by students to be equally as rigorous?

First, multiple raters assessed problem solving for dozens of students and 19 different problems. Next, an established survey tools measured constructs of problem solving and perception of problem difficulty, which addressed the second research question.

3.1 Problem-Solving Ability

The scores students received on the PROCESS instrument were calibrated using the Rasch model. The Rasch computed PROCESS scores were set to 0–18 to mirror the raw score range, with 0 indicating the lowest level of problem-solving ability and 18 representing the highest. A series of ANCOVA tests showed that at the beginning of the respective courses, the treatment and control groups were similar in chemical engineering problem solving ability as measured by PROCESS scores, when controlling for different level of math course completed the two groups (Table 4). The control group did exhibit statistically significantly higher PROCESS scores on TB 6 and 8; further discussion would focus on specific course details, which is outside the scope of this paper. Problems at the end of the course (TB 9 and 10) revealed that the difference in group performance on Textbook problems diminished with the treatment group narrowing the gap in performance and scoring higher on average for TB 9, which will be discussed later.

Analysis of the treatment group performance by problem type revealed a statistically significantly better performance on the YouTube problems. The treatment group scored higher on YouTube problems (13.2 ± 2.6) than Textbook problems (12.4 ± 2.9), t (632) = 3.6, p = 0.001 (Fig. 2). Videos provided alongside with problem statements may be responsible for why students displayed higher problem-solving acumen by helping students to visualize and understand better [11, 17]..

Examining the PROCESS scores by item revealed similarities between the Textbook and YouTube problems (Table 5). Rankings from 1 (most difficult) to 6 (easiest) revealed the relative

Table 4. Comparison of Rasch estimated PROCESS scores on Textbook problems

Problem	Control	Treatment	*p*
TB1	15.6 ± 1.0	13.5 ± 0.8	0.15
TB2	14.5 ± 1.0	11.4 ± 1.1	0.078
TB3	12.1 ± 0.7	10.3 ± 0.7	0.13
TB4	14.0 ± 0.7	12.3 ± 0.5	0.10
TB5	13.8 ± 0.8	13.6 ± 0.6	0.81
TB6	**13.9** ± 0.7	**10.8** ± 0.6	**0.008***
TB7	14.7 ± 0.7	13.1 ± 0.6	0.15
TB8	**16.8** ± 0.7	**14.3** ± 0.6	**0.029***
TB9	12.6 ± 0.8	12.7 ± 2.2	0.88
TB10	12.5 ± 0.82	11.9 ± 1.8	0.56

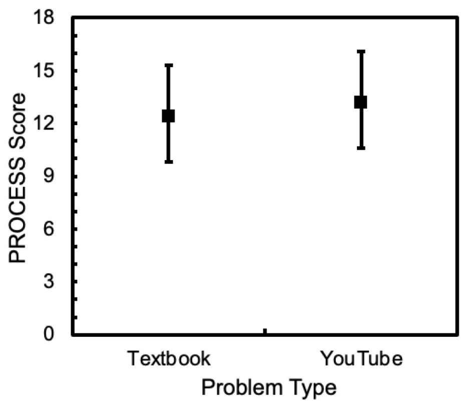

Fig. 2. Treatment group performance by problem type estimated from 19 problems. Error bars are standard deviations.

Table 5. Rank order of PROCESS item difficulty by problem type

Item	Textbook	YouTube
P (Identify Problem)	5.3 ± 0.6	5.2 ± 0.4
R (Represent)	4.6 ± 1.6	5.7 ± 0.7
O (Organize)	3.4 ± 0.5	3.0 ± 0.0
C (Calculate)	2.2 ± 0.4	2.0 ± 0.0
S (Solution Completion)	4.3 ± 0.9	4.1 ± 0.3
S (Solution Accuracy)	1.0 ± 0.0	1.0 ± 0.0

Table 6. Comparison of Rasch estimated NASA TLX scores for Textbook problems

Problem	Control	Treatment	*p*
TB1	48 ± 11	46 ± 10	0.35
TB2	52 ± 9	49 ± 10	0.18
TB3	61 ± 8	57± 12	0.25
TB4	50 ± 11	49 ± 10	0.59
TB5	57 ± 8	56 ± 11	0.68
TB6	58 ± 8	58 ± 9	0.94
TB7	**59** ± 10	**50** ± 10	**0.001***
TB8	47± 13	41 ± 14	0.12
TB9	57± 8	52 ± 11	0.22
TB10	66 ± 11	58 ± 11	0.06

Table 7. Rank order of NASA TLX item difficulty by problem type

Task	Textbook	YouTube
Mental Demand	1.7 ± 0.5	1.9 ± 0.3
Physical Demand	4.8 ± 0.9	4.4 ± 0.5
Temporal Demand	5.4 ± 0.7	5.3 ± 0.7
Performance	4.8 ± 0.7	5.2 ± 0.9
Effort	1.3 ± 0.5	1.1 ± 0.3
Frustration	3.0 ± 0.0	3.0 ± 0.0

difficulty of each PROCESS item for each homework problem type. The relative order of difficulty agreed between homework problem types with the exception of flipping the order of the two easiest items (Identify problem and Represent). Visuals included in YouTube problems may be the reason why representing a problem through process flow diagrams appear to be the easiest task during problem solving. Solution accuracy proved to be the most difficult item across all problems, which agrees with intuition that the final step in problem solving contained the most errors. Within the Textbook problems, no variation was found in item difficulty between treatment and control groups. The order of PROCESS item difficulty fluctuated more (quantified by standard deviations in Table 5) among the Textbook problems compared to the YouTube problems.

3.2 Perception of Problem Difficulty

Responses to the NASA TLX quantified perceived problem difficultly and scores were calibrated with the Rasch model. A range from 0 to 100 mimics the typical NASA TLX range, with 0 being the lowest level of perceived difficulty. Student's t-tests revealed that the control group and the treatment group perceived the Textbook problems to be of equal rigor with one exception – problem TB 7 (Table 6). Although the treatment group perceived nine out of ten Textbook problems to be slightly less rigorous than the control group, only TB 7 was perceived to be statistically significantly easier. TB 7 was a two-component flash separation problem involving a multiphase system and vapor-liquid equilibrium. Aggregating all problems by type for the treatment group found no statistically significant difference. The treatment group perceived YouTube (52 ± 12) and Textbook (51 ± 12) problems to be of similar rigor, $t\ (1088) = 1.6$, $p = 0.11$(see Fig. A.6).

Similar to the analysis of the PROCESS rubric, ranking the NASA TLX scores by item revealed similarities between the Textbook and YouTube problems overall (Table 7). The relative order of difficulty (where 1 indicated the task students found most difficult about the problem) stayed steady between problem types. Perceived effort, i.e., how hard the students had to work to accomplish their level of performance, was the task students identified as the most difficult for both problem types. Frustration level remained constant between problem types, with students indicating moderate levels of insecurity and stress when solving the problems. The high ranking of mental demand when solving engineering problems was expected compared to physical or temporal demand. More specifically, temporal demand and physical demands contributed least to problem difficulty owing to the fact that sufficient time, usually 1 week, was allowed for students to complete problems.

3.3 Relationship Between Problem-Solving Ability and Problem Difficulty

A significant linear regression equation was found between perception of YouTube problem difficulty and problem-solving ability for the treatment group ($F(1, 7) = 9.6$, $p = 0.017$) (see Table A.7 for detailed scores). A similar significant linear regression equation was found for Textbook problems ($F(1, 8) = 7.1$, $p = 0.034$) (see Table A.8 for detailed scores). A strong, negative correlation between YouTube problem difficulty and problem-solving ability ($R = -0.76$) similar to the same correlation for Textbook problems ($R = -0.67$) (Fig. 3). Yet, examining the slope of each regression line revealed that the treatment group would be expected to achieve higher scores on the YouTube problems in spite of higher levels of perceived difficulty, when compared to solving Textbook problems. This relationship found that a ten-point increase on the NASA TLX (i.e., a 10% increase in perceived problem difficulty) had different implications depending on which problem type the treatment group solved. For treatment group students completing Textbook problems, a 10% increase in perceived problem difficulty would entail a predicted decrease in PROCESS score of 7.2% and only 6.1% PROCESS score decrease when solving YouTube problems.

Significant relationships were found between treatment and control group problem-solving ability and perceived difficulty of Textbook problems. A significant linear regression was found between perception of Textbook problem difficulty and problem-solving ability for the control group, ($F(1, 8) = 11.0$, $p = 0.01$) (see Table A.9 for detailed scores). A similar although less robust relationship was found for Textbook problems and the treatment group, ($F(1, 8) = 6.5$, $p = 0.034$). Overall, the treatment group expressed lower levels of perceived difficulty and problem-solving ability on the Textbook problems compared to the control group (see Fig. A.10).

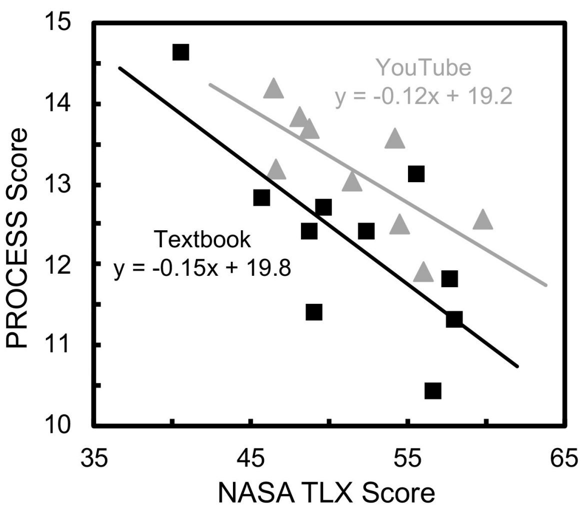

Fig. 3. Treatment group relationship between PROCESS and NASA TLX scores for YouTube (triangles) and Textbook (squares) problems.

A measure of relative problem-solving ability and perception of difficulty was computed for each group at the beginning and the end of the course by averaging the Rasch-calibrated PROCESS scores and NASA TLX ratings for the two Textbook problems at the beginning (TB1 and TB2) and two near the end of the course (TB9 and TB10). The first set of problems (TB1 and TB2) covered concepts of volume percent/mole ratio calculation and basic mass balances for non-reacting systems. The later problems (TB9 and TB10) covered both material and energy balance concepts for reacting systems. At the beginning of course, control group significantly displayed higher problem-solving ability than treatment group ($p = 0.001$). However, towards the end of course, a convergence of PROCESS scores between control and treatment groups revealed similar problem-solving abilities. Whereas the control group experienced significantly lower (~13%) PROCESS scores between beginning and end, lower PROCESS scores in treatment group problem solving at the end of the semester compared to beginning (<1%) were negligible ($p = 0.897$). An increase in perceptions of problem difficulty over the course of the semester, as measured by TLX scores, corresponded to the decrease in PROCESS scores. Both control and treatment groups found the beginning Textbook problems to be equally rigorous (TLX average scores of 49.9 and 47.3, respectively) and significantly easier than Textbook problems at the end of the course (TLX scores of 60.7 and 55.0, respectively). The significantly higher perception of problem difficulty displayed by the control group at the end of the semester ($t(96) = 2.01$, $p = 0.047$) may in part explain the lower PROCESS scores compared to the treatment group (Fig. 4). The same relative effects were found when considering Rasch-calibrated PROCESS scores on their own as well as when taking highest level of math into consideration as a covariate. These data suggest that YouTube problems may be beneficial in that the initially lower scoring treatment group gained sufficient problem-solving skills to eliminate the gap observed early in the course. Overall, by the end of the semester, the treatment group increased their chemical engineering problem solving ability as measured by PROCESS scores. Additional scoring and a second annual cohort were collected to answer these questions more clearly in future work.

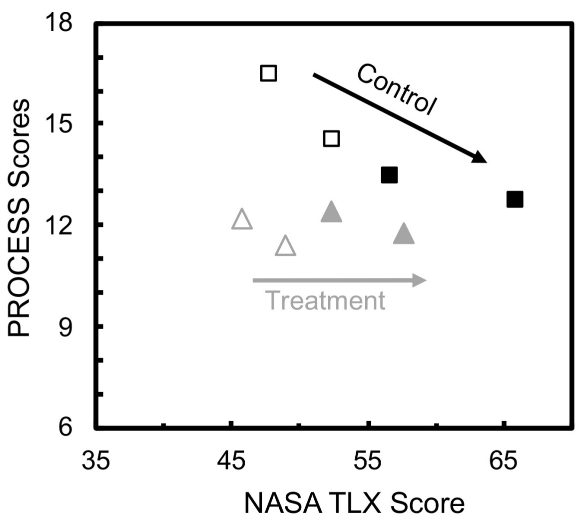

Fig. 4. Pre/post PROCESS scores and NASA TLX scores for treatment (triangles) and control (squares) group across semester. Open symbols represent scores for two Textbook problems completed early in the study while filled symbols are scores for two Textbook problems at the end of the study. Arrows are to guide the eye.

4. Conclusion

Homework-style, YouTube-inspired problems have been implemented in an undergraduate course in material and energy balances. YouTube problems were utilized as alternative Textbook homework problems for students and covered a wide variety of topics in material and energy balance course. A set of 9 YouTube problems in combination with 10 Textbook problems served as the basis for examining problem-solving ability and perception of rigor. Through implementation of pseudo-control/treatment design, research examined impacts of replacing Textbook problems with YouTube problems. Research questions were directed towards evaluating rigor and problem solving utilizing both evidence-based strategies and surveys to measure parameters associated to learning.

NASA TLX survey measured difficulty of problems across six items. Overall analysis found similar perception of problem rigor between YouTube and Textbook problems in responses for both treatment and control groups. Item analysis identified mental demand, effort, and frustration as the most significant factors to problem difficulty in solving material and energy balance problems.

An established problem-solving rubric called PROCESS was revised and implemented across problem types and groups. YouTube problems may be beneficial in that the lower scoring treatment group gained sufficient problem-solving skills to eliminate the gap observed early in the course. Inclusion of videos alongside problem statement might be responsible for higher problem-solving ability displayed by students when solving YouTube problems. However, one limitation of this study was that video view rates for students solving the YouTube problems were not quantified.

Item analysis within PROCESS identified solution accuracy stage as the most difficult item within PROCESS which is not surprising since solution accuracy measures the final outcome of problem solving and low scores might be compounding from missing or incorrect steps identified with earlier stages of problem solving, such as Organization and Calculations components. Therefore, addressing challenges with earlier stages of problem solving may improve Solution accuracy. Overall, PROCESS could serve as a feedback tool for instructors allowing them to identify and address stages of problem solving where students are most challenged.

Problem-solving skills indicated by PROCESS scores correlated negatively with perception of problem difficulty from NASA TLX. Students exhibited better problem-solving skills on problems perceived to be less demanding. Interestingly, perception of problem difficulty correlated more weakly with problem-solving ability for YouTube problems compared to Textbook problems. A weaker correlation of problem difficulty with problem-solving skills may have resulted from the incorporation of videos into YouTube problems enabling students to visualize and aid the problem-solving process.

Between the four raters, 19 different problems, over 2,600 PROCESS scores, and 1,200 TLX surveys were analyzed for this paper. Obtaining a large set of PROCESS scores was very labor intensive with every solution was assessed by each rater. In the future, a more streamlined scoring plan using inter-rater reliability (as described in [48]) could cut down the number of solutions scored and time required to execute similar research. Alternatively, a recent qualitative study measured improved learning attitudes for students who solved YouTube problems [61]. A future study will deploy the same experimental strategy on an additional cohort, which will hopefully generalize some of the findings presented here.

Acknowledgements – Charlene Czerniak, Lisa Young, Katherine Roach, and Caleb Sims have contributed at various stages of the research. This material is based upon work supported by the National Science Foundation under Grant No. DUE 1712186. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work was completed within the framework of University of Toledo IRB protocol 202214.

References

1. Internet Usage Statistics; The Internet Big Picture: World Internet Users and 2018 Population Stats, https://www.internetworldstats.com/stats.htm, Accessed January 2019.
2. G. E. Kennedy, T. S. Judd, A. Churchward, K. Gray and K.-L. Krause, First year students' experiences with technology: Are they really digital natives? *Australasian Journal of Educational Technology*, **24**(1), 2008.
3. How Net Generation Students Work, https://people.howstuffworks.com/how-net-generation-students-work.htm, Accessed March 2021.
4. J. M. Lang, *Small Teaching: Everyday Lessons from the Science of Learning*, John Wiley & Sons. 2016.
5. M. W. Liberatore, Improved student achievement using personalized online homework for a course in material and energy balances, *Chemical Engineering Education*, **45**(3), pp. 184–190, 2011.
6. The Future of Jobs: employment, Skills and Work force Strategy for the Fourth Industrial Revolution, http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf, Accessed May 2021.
7. Top 10 Jobs in 2030: Skills You Need Now to Land the Jobs of the Future: Future Skills, https://www.crimsoneducation.org/us/blog/jobs-of-the-future, Accessed January 2019.
8. C. M.Vest, *Infusing Real World Experiences into Engineering Education*, National Academies Press. 2012.
9. G. T. Daigger, L. A. Hammon, R. M. Haynes, L. H. Jamieson and D. N. Wormley, *Infusing real world experiences into engineering education*, 2012.
10. J. Hawkins and S. Blakeslee, *On intelligence*, St. Martin's Griffin, 2005.
11. R. Tasker and R. Dalton, Research into practice: visualisation of the molecular world using animations, *Chemistry Education Research and Practice*, **7**(2), pp. 141–159, 2006.
12. F. Wen and E. Khera, Identify Solve Broadcast your own transport phenomena: Student created YouTube videos to foster active learning in mass and heat transfer, *Chemical Engineering Education*, **50**(3), pp. 186–192, 2016.
13. A. Velazquez-Marcano, V. M. Williamson, G. Ashkenazi, R. Tasker and K. C. Williamson, The Use of Video Demonstrations and Particulate Animation in General Chemistry, *Journal of Science Education and Technology*, **13**(3), pp. 315–324, 2004.
14. M. W. Liberatore, D. W. M. Marr, A. M. Herring and J. D. Way, Student-created homework problems based on YouTube videos. *Chemical Engineering Education*, **47**(2), pp. 122–132, 2013.
15. J. Medina, *Brain Rules 12 Principles for Surviving and Thriving at Work, Home, and School*, Pear Press. 2008.
16. M. B. McGrath and J. R. Brown, Visual learning for science and engineering, *IEEE Comput Graph Appl*, **25**(5), pp. 56–63, 2005.
17. S. M. Al-Balushi and S. H. Al-Hajri, Associating animations with concrete models to enhance students' comprehension of different visual representations in organic chemistry, *Chem. Educ. Res. Pract.*, **15**(1), pp. 47–58, 2014.
18. N. Doidge, *The brain that changes itself: Stories of personal triumph from the frontiers of brain science*, Penguin. 2007.
19. N. Doidge, *The brain's way of healing: Remarkable discoveries and recoveries from the frontiers of neuroplasticity*, Penguin Books. 2016.
20. S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proceedings of the National Academy of Sciences*, **111**(23), pp. 8410–8415, 2014.
21. J. L. Falconer, Why not try active learning? *Aiche Journal*, **62**(12), pp. 4174–4181, 2016.
22. S. R. Cavanagh, *The Spark of Learning: Energizing the College Classroom with the Science of Emotion*. West Virginia University Press. 2016.
23. M. W. Liberatore, C. R. Vestal and A. M. Herring, YouTube Fridays: Student led development of engineering estimate problems, *Advances in Engineering Education*, **3**(1), pp. 1–16, 2012.
24. M. W. Liberatore, R. M. Morrish and C. R. Vestal, Effectiveness of Just in Time Teaching on Student Achievement in an Introductory Thermodynamics Course, *Advances in Engineering Education*, **6**(1), pp. n1, 2017.
25. M. W. Liberatore, Active Learning and Just-in-time Teaching In a Material and Energy Balances Course, *Chemical Engineering Education*, **47**(3), pp. 154–160, 2013.
26. M. W. Liberatore, High textbook reading rates when using an interactive textbook for a Material and Energy Balances course, *Chemical Engineering Education*, **51**(3), pp. 109–118, 2017.
27. M. W. Liberatore, YouTube Fridays: Engaging the Net Generation in 5 Minutes a Week, *Chemical Engineering Education*, **44**(3), pp. 215–221, 2010.
28. K. R. Hadley and K. A. Debelak, Wiki technology as a design tool for a capstone design course, *Chemical Engineering Education*, **43**(3), pp. 194–200, 2009.
29. T. F. Edgar, Information Technology and ChE Education: Evolution or Revolution? *Chemical Engineering Education*, (Fall): pp. 290–295, 2000.
30. M. Liberatore, U. Asogwa, A. Malefyt, C. Czerniak, G. Mentzer and T. Duckett. Student-Generated Problems that Reverse Engineer YouTube Videos, *ASEE Annual Conference Proceedings*. 2018.
31. M. T. Chi, Active-constructive-interactive: a conceptual framework for differentiating learning activities, *Topics in Cognitive Science*, **1**(1), pp. 73–105, 2009.
32. C. Zhou, A. Kolmos and J. F. D. Nielsen, A problem and project-based learning (PBL) approach to motivate group creativity in engineering education, *International Journal of Engineering Education*, **28**(1), pp. 3–16, 2012.
33. J. L. Docktor, J. Dornfeld, E. Frodermann, K. Heller, L. Hsu, K. A. Jackson, A. Mason, Q. X. Ryan and J. Yang, Assessing student written problem solutions: A problem-solving rubric with application to introductory physics, *Physical Review Physics Education Research*, **12**(1), pp. 010130, 2016.
34. K. Montgomery, Authentic tasks and rubrics: Going beyond traditional assessments in college teaching, *College Teaching*, **50**(1), pp. 34–40, 2002.
35. L. Benson, L. Grigg, D. R. Bowman, M. Cook and R. P. Pargas, CU thinking problem solving strategies revealed, *ASEE Annual Meeting*, Vancouver, BC, p. 1576, 2011.
36. S. J. Grigg and L. Benson, Effects of Student Strategies on Successful Problem Solving, *ASEE Annual Meeting*, San Antonio, TX, p. 21266, 2012.

37. S. J. Grigg and L. Benson, Promoting problem solving proficiency in first year engineering process assessment, *ASEE Annual Meeting*. Seattle, WA, 2015.
38. S. J. Grigg and L. C. Benson, A coding scheme for analysing problem-solving processes of first-year engineering students, *European Journal of Engineering Education*, **39**(6), pp. 617–635, 2014.
39. S. J. Grigg, J. Van Dyken, L. Benson and B. Morkos, Process analysis as a feedback tool for development of engineering problem solving skills, *ASEE Annual Meeting*, Atlanta, p. 6505, 2013.
40. S. J. Grigg, S. K. Garrett and L. Benson, Using the NASA-TLX to Assess First Year Engineering Problem Difficulty, *IIE Annual Conference*, pp. 1–10, 2012.
41. S. K. Luger and J. Bowles, Two methods for measuring question difficulty and discrimination in incomplete crowdsourced data, *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.
42. M. A. Vidulich and P. S. Tsang, Assessing subjective workload assessment: a comparison of SWAT and the NASA-bipolar methods, *Proceedings of the Human Factors Society Annual Meeting*, pp. 71–75, 1985.
43. G. B. Reid and T. E. Nygren, *The subjective workload assessment technique: A scaling procedure for measuring mental workload*, in *Advances in psychology*, Elsevier, pp. 185–218, 1988.
44. K. E. Cook, Y.-L. Han, T. R. Shuman and G. Mason, Effects of integrating authentic engineering problem centered learning on student problem solving, *International Journal of Engineering Education*, **33**(1), pp. 272–282, 2017.
45. B. S. Bloom, ed, *Taxonomy of Education Objectives: the classification of educational goals*, Longsmans, Green, New York, pp. 1956.
46. R. H. Bruning, G. J. Schraw and R. R. Ronning, *Cognitive psychology and instruction*, Merrill, Upper Saddle River, N.J., 1999.
47. U. Asogwa, M. W. Liberatore and A. P. Malefyt, Problem Solving When Using Student-Written YouTube Problems, *ASEE Annual Conference & Exposition*, 2019.
48. T. R. Duckett, U. Asogwa, M. Liberatore and G. Mentzer, Assessing the reliability of a problem solving rubric when using multiple raters, *ASEE Annual Conference Proceedings*, 2019.
49. G. Engelhard Jr and S. Wind, *Invariant measurement with raters and rating scales: Rasch models for rater-mediated assessments*, Routledge, 2017.
50. U. Boser, *Learn Better: Mastering the Skills for Success in Life, Business, and School, or, How to Become an Expert in Just About Anything*, Rodale Books. p. 304, 2017.
51. R. A. Grier, How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **59**(1), pp. 1727–1731, 2015.
52. S. G. Hart and L. E. Staveland, *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, in *Human Mental Workload*, pp. 139–183, 1988.
53. D. Sharek, A Useable, Online NASA-TLX Tool, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **55**(1), pp. 1375–1379, 2011.
54. K. C. Hendy, K. M. Hamilton and L. N. Landry, Measuring subjective workload: when is one scale better than many? *Human Factors*, **35**(4), pp. 579–601, 1993.
55. L. M. Lozano, E. García-Cueto and J. Muñiz, Effect of the number of response categories on the reliability and validity of rating scales, *Methodology*, **4**(2), pp. 73–79, 2008.
56. D. Andrich, A rating formulation for ordered response categories, *Psychometrika*, **43**(4), pp. 561–573, 1978.
57. J. M. Linacre, WINSTEPS Rasch measurement computer program, 2006.
58. T. G. Bond and C. M. Fox, *Applying the Rasch model: Fundamental measurement in the human sciences*, Psychology Press, 2001.
59. B. D. Wright and G. N. Masters, *Rating scale analysis*, MESA Press, 1982.
60. S. Weisberg, *Applied linear regression*, **528**, John Wiley & Sons, 2005.
61. U. Asogwa, T. R. Duckett, G. Mentzer and M. W. Liberatore, Impact of YouTube Homework Problems on Students' Learning Attitudes, *Chemical Engineering Education*, 2021.

Uchenna Asogwa is a PhD student in Chemical Engineering at the University of Toledo. His current research involves reverse engineering online videos, the rheology of complex fluids, and fuel cell membranes.

T. Ryan Duckett is a research associate with Acumen Research and Evaluation, LLC., a program evaluation and grant writing company that specializes in STEM and early childhood education. He is a PhD student in the Research and Measurement department at the University of Toledo.

Amanda Malefyt is currently Chair and Associate Professor in the McKetta Department of Chemical and Bioprocess Engineering at Trine University. She received her PhD from Michigan State University and bachelor's degree from Trine (formerly Tri-State) University. Her research interests include engineering education and nucleic acid therapeutics.

Lindsey Stevens is a Middle School STEM teacher at Maumee Valley Country Day School with experience teaching High School Biology at St. Ursula Academy. Her current research involves creating reverse engineering online videos and scoring assessments for comprehension.

Gale A. Mentzer, the owner and director of Acumen Research and Evaluation, LLC, has been a professional program evaluator since 1998. She holds a PhD in Educational Research and Measurement from The University of Toledo and a Master of Arts in English Literature and Language – a unique combination of specializations that melds quantitative and qualitative methodologies. She and has extensive experience in the evaluation of projects focused on STEM education including evaluations of several multi-million dollar federally funded projects. Previously she taught graduate level courses for the College of Education at The University of Toledo in Statistics, Testing and Grading, Research Design, and Program Evaluation.

Matthew W. Liberatore is a Professor of Chemical Engineering at the University of Toledo. He earned a B.S. degree from the University of Illinois at Chicago and MS and PhD degrees from the University of Illinois at Urbana-Champaign, all in chemical engineering. His current research involves the rheology of complex fluids as well as active learning, reverse engineering online videos, and interactive textbooks. His website is: http://www.utoledo.edu/engineering/chemical-engineering/liberatore/

Appendix

Table A.1. Titles for YouTube and Textbook problems assigned and code names

Textbook Problems Title	Code	YouTube Problem Title	Code
Volume percent to mole fractions for a ternary mixture	TB1	Biodiesel Production	YT1
Mixing chamber with Isopropanol	TB2	Nylon production	YT2
Algae Processing	TB3	Glass making	YT3
Methanol reactor	TB4	100 Tons of Dynamite	YT4
Ethane reaction with purge	TB5	CO2 Hydrogenation to Methanol	YT5
Isopropanol mixing	TB6	Carbon Dioxide reacted to form Pure Methanol	YT6
Benzene aniline flash	TB7	Artificial Trees That Absorb CO2	YT7
Rainy reservoir	TB8	How Does a Dehumidifier Work?	YT8
Ice ammonia heat exchanger	TB9	Ethanol Production Process	YT9
Ethane combustion energy balance	TB10		

Table A.2. Modified PROCESS rubric for problem solving using handwritten solutions

Problem Solving Process/ Category	Explicit Tasks Performed	Level of Completion				
		Missing 0 points	Inadequate 1 point	Adequate 2 points	Accurate 3 points	Error sources
Identify Problem and System Constraints	Identified unknown	Did not explicitly identify or define the problem/system	Completed few problem/system definition tasks with many errors	Completed most problem/system definition tasks with few errors	Clearly and correctly identified and defined the problem/system	Identified Incorrect unknown
						Identified fewer unknown than required
Represent the Problem	Drew a flow diagram	No representation drawn, no relationships indicated	Drew a representation or related variables, but not both	Drew a representation and related most variables with some errors	Drew a representation with all streams and process units and indicated variable relationships correctly	Too many/ fewer streams than required
						Wrong location of process unit
	Labeled the flow diagram					Wrong location of stream/process unit
Organize Knowledge	Identified known values	Did not explicitly organize information about the problem	Completed few information organization tasks	Completed most information organization tasks	Fully organized information needed to solve the problem	Solved using wrong values of known values
	Identified equations (atomic or component mass/mole balance equations)					Missing term in balances
	Identified extra equations example, Conversion, percentage excess, recycle ratio					Missing term in extra equations
	Identified other useful equations example, Antoine, Raoult's law equations					Wrote Incorrect formula
Calculation (Allocate Resources)	Manipulated/ solved equations	No work shown	Partially documented execution tasks (Work showed some evidence of relevant tasks)	well documented execution tasks but with few omissions	Fully documented execution tasks (Work showed evidence of relevant tasks)	Calculation error from extra equation
	converted to the required units(optional)					Did not simplify equations correctly
						Calculation error from balance
Final Solution Completion	Provided answers to the problem statement	All answers are missing	Explicitly evaluated some of the solution or evaluated incorrect unknown	Explicitly evaluated most of the solution	Evaluated all the solutions required	
Final Solution Accuracy	Correct/incorrect values for answers to all parts of the problem.	Missing Answer	Provides mostly incorrect answers or no units to all parts of the multipart problem answer	Provides mostly correct answers and units to all parts of the multipart problem	Provides correct answers and units to all parts of the multipart problem	presents wrong answers
	Correct units for answers to all parts of the problem.					wrong units
TOTAL SCORE						

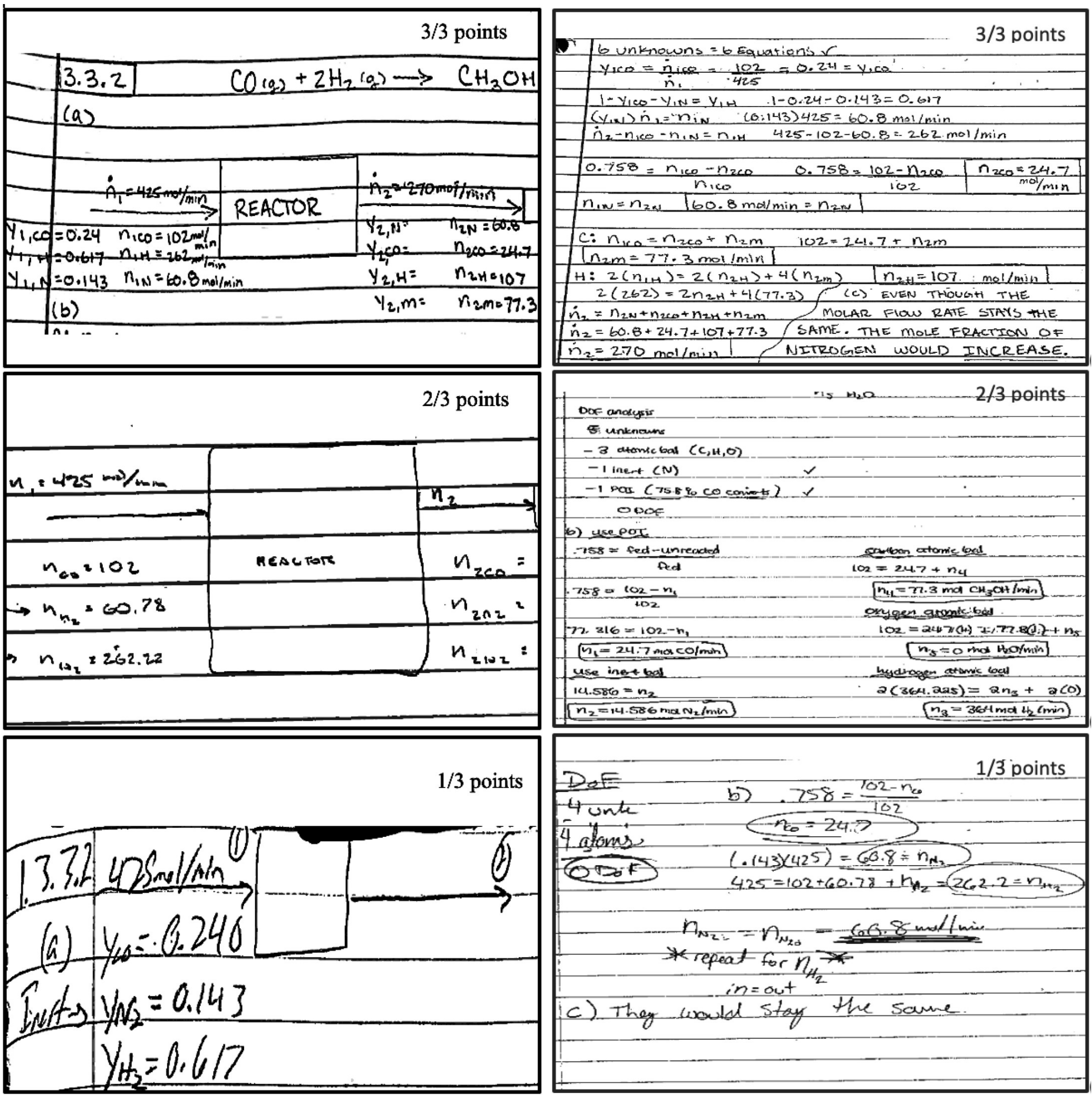

Fig. A.3. Samples of student written solutions for methanol reactor problem in Fig. 1 to illustrate how problem representation stage (R) in the left and calculation stage (C) in the right of PROCESS rubric is applied.

The proficiency required to earn 3 points in the R section of the rubric include drawing and labelling correctly a process flow diagram that accurately represents the system (see Fig. 1). To earn 3 points, a solution should contain a sketch of the reactor with inlet stream containing CO, H_2 and N_2 and an outlet stream CO, H, N_2, and CH_3OH. When student's work contained most, but not all components in the streams, the solution earned 2 points. 1 point is assigned when an attempt was made to draw the process diagram with few details.

Achieving 3 points in the C section required detailed calculations leading to a solution for all parts of the problem. A solution was scored 2 points when most of the calculations leading to the solution for all parts were completed for example when solution failed to show how molar flow rate of H_2 in the inlet stream was obtained. Finally, a solution was scored 1 point when student work showed few calculations leading for multiple parts problems as seen in example where solution failed to show work for obtaining molar flow rate for H_2 and CH_3OH in the outlet stream.

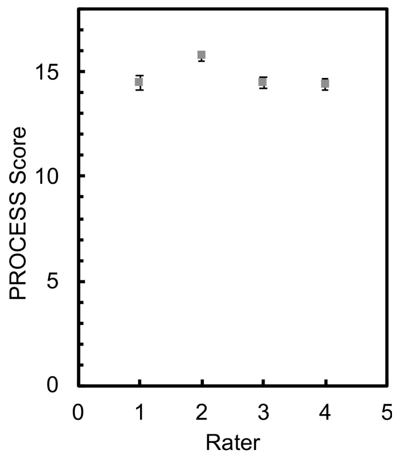

Fig. A.4. Average rater score across PROCESS ratings with 95% CI.

Table A.5. NASA Task Load Index question used by students to rate problem rigor

TLX Questions
How mentally demanding was the task?
How physically demanding was the task?
How hurried or rushed was the pace of the task?
How successful were you in accomplishing the task?
How hard did you have to work to accomplish your level of performance?
How insecure, discouraged, irritated, stressed, or annoyed were you?

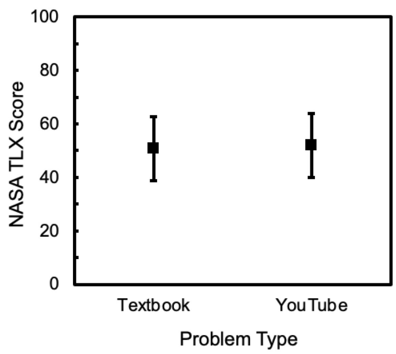

Fig. A.6. Overall NASA TLX scores for Treatment group when completing YouTube and Textbook problems.

Table A.7. Rasch estimated PROCESS and NASA TLX scores on YouTube problems for Treatment group

YouTube Problems	Treatment	
	NASA TLX	PROCESS
YT 1	56 ± 11	11.9 ± 3.6
YT 2	60 ± 10	12.6 ± 2.4
YT 3	54 ± 12	12.5 ± 2.6
YT 4	51 ± 13	13.0 ± 2.3
YT 5	46 ± 12	14.2 ± 2.3
YT 6	49 ± 12	13.7 ± 2.1
YT 7	47 ± 11	13.2 ± 1.5
YT 8	48 ± 11	13.8 ± 2.6
YT 9	54 ± 10	13.6 ± 2.8

Table A.8. Rasch estimated PROCESS and NASA TLX scores on Textbook problems for Treatment group

Textbook problems	Treatment	
	NASA TLX	PROCESS
TB 1	46 ± 10	12.8 ± 5.0
TB 2	49 ± 10	11.4 ± 4.8
TB 3	57± 12	10.4 ± 2.6
TB 4	49 ± 10	12.4 ± 2.7
TB 5	56 ± 11	13.1 ± 2.4
TB 6	58 ± 9	11.3 ± 3.0
TB 7	50 ± 10	12.7 ± 2.4
TB 8	41 ± 14	14.6 ± 2.6
TB 9	52 ± 11	12.4 ± 2.2
TB 10	58 ± 11	11.8 ± 1.8

Table A.9. Rasch estimated PROCESS and NASA TLX scores on Textbook problems for control group

Textbook problems	Control	
	NASA TLX	PROCESS
TB 1	48 ± 11	16.4 ± 1.2
TB 2	52 ± 9	14.5 ± 2.2
TB 3	61 ± 8	12.1 ± 1.9
TB 4	50 ± 11	13.9 ± 1.8
TB 5	57 ± 8	14.5 ± 3.0
TB 6	58 ± 8	13.3 ± 1.8
TB 7	59 ± 10	15.2 ± 2.5
TB 8	47± 13	16.4 ± 2.5
TB 9	57± 8	13.4 ± 2.9

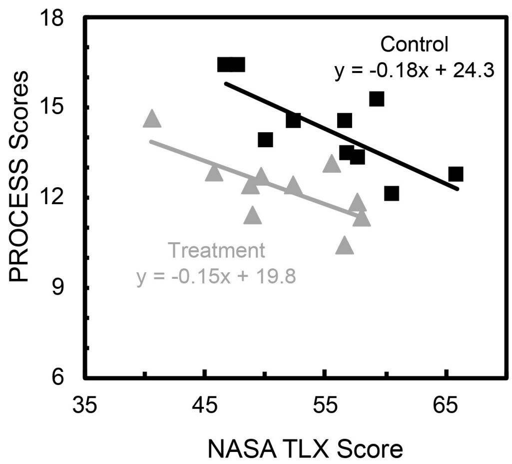

Fig. A.10. Relationship between PROCESS and NASA TLX scores when control (squares) and treatment (triangles) group solve Textbook problems.