A Thesis

entitled

Developing Problem Solving Skills using Student Generated Problems that Reverse Engineer YouTube Videos

by

Uchenna Asogwa

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Master of Science Degree in

Chemical Engineering

____

Dr. Matthew W Liberatore, Committee Chair

____

Dr. Yakov Lapitsky, Committee Member

____

Dr. Carmen Cioc, Committee Member

____

Dr. Amanda C. Bryant-Friedrich, Dean
College of Graduate Studies

The University of Toledo
August 2020

An Abstract of

Developing Problem Solving Skills using Student Generated Problems that Reverse Engineer YouTube Videos

by

Uchenna Asogwa

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Master of Science Degree in Chemical Engineering

The University of Toledo
August 2020

Problem solving is a signature skill of engineers. Here, problem solving is employed when students apply course concepts to reverse engineer YouTube videos and solve new student-written, homework-style problems (YouTube problems). Replacing textbook problems with YouTube problems, this research focuses on examining the rigor of YouTube problems as well as students' problem-solving skills on textbook and YouTube problems. A quasi-experimental, treatment/control group design was employed, and data was collected and evaluated using multiple instruments. First, rigor of homework problems was examined using the NASA Task Load Index. Also, problem solving was assessed using a previously-developed rubric called PROCESS Problem definition, Representing the problem, Organizing the information, Calculations, Evaluating the solution, Solution communication, and Self-assessment. PROCESS was modified to independently measure completeness and accuracy of student responses, as well as identify errors committed in material and energy balances. In the treatment group, students were assigned ten textbook problems and nine YouTube problems. In addition to obtaining an evidence-based assessment of problem solving via PROCESS, students' learning attitudes, overall and with respect to problem solving, were measured via a self-reported survey known as Colorado Learning Attitudes about Science Survey (CLASS). Utilizing YouTube problems in classroom did not influence learning attitudes of students negatively. Students reported that YouTube problems possessed similar rigor as Textbook problems. Instead, students solving YouTube problems measured small effect size improvement in problem-solving skills.

Acknowledgements

I would like to appreciate my advisor, Dr. M. W. Liberatore, for his support and guidance throughout this research. Without his help, this project and thesis would not have been possible. I also thank students who have worked on this project, including Lisa Young, Katherine Roach, Caleb Sims and contributed significantly at various stages of the research.

I would like to thank all the collaborators involved in this project, Amanda, Gale and Ryan for their various insights in this research. I appreciate my family for their love and support throughout this 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.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Background

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 utilization of electronic devices, such as cell phones, tablets, and computers. These students are often referred to as digital natives [2]. Near 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 differences are being identified between digital natives and past generations. In many cases, digital natives show a preference for visual media compared to textual modes of learning, 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. Recent efforts to incorporate the Accrediting Board for Engineering and Technology (ABET) standards to emphasize problem solving and knowledge of current issues have found that infusing real world situations into engineering education helps students' understanding become more integrated [8, 9]. Therefore, tying engineering problem solving 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 trumps other senses in creating both short term and long-term learning [10]. Visual representation is an important part of successfully solving complex problems. Visual learning methods open new ways of problem solving, 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 [20-22]. Also, involving students' enthusiasm is advantageous to learning [23]. 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, 24-28]. Implementation of technology as a form of active learning is a useful approach that connects students and learning [29, 30]. Therefore, engaging the current generation of students using technology, like YouTube, in a positive way is one motivation directing this project.

1.1 What is YouTube Pedagogy

Our YouTube pedagogy started as a way to introduce and engage students in thermodynamics and material and energy balance courses. Originally called YouTube Fridays, the first five minutes of Friday class sessions were dedicated to course related videos selected by a group of students. At the end of the semester, surveys showed that the vast majority (> 80%) of students felt they had better understanding of the field of chemical engineering from participating in YouTube Fridays [28]. In subsequent semesters, students selected YouTube videos and created engineering estimate problems related to the course material. Surveys showed that the majority of the students felt they had a better understanding of the course topic of thermodynamics, could relate thermodynamics to real world phenomena, and felt confident solving engineering estimate problems. The vast majority of survey feedback about YouTube Fridays was positive and provided students a mechanism to apply classroom concepts to open-ended, real world situations [24]. While open-ended problems provide good practice of engineering fundamentals, replacing closed, dated textbook problems represented the next evolution.

The YouTube pedagogy provides an alternate to professor-centric lectures, screencasts, and textbook problems. Here, students are not required to create videos and rarely do, which agrees with findings related to watching videos or generating a single video leading to little or no improvement of students' conceptual learning [31]. Videos continue to be taken from YouTube or other websites in the public domain.

Over 400 student-written YouTube problems have been created in recent years [14, 24, 32]. While the writing is largely open ended, a small number of boundaries keep the students' authoring focused. The assignment is initiated by students watching YouTube videos and selecting one 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.

1.2 Example YouTube Problems

Problems that can be implemented in class, as part of homework sets, or in quizzes/exams has been developed with the YouTube pedagogy. YouTube problems possess features that examine student's learning at numerous levels of Bloom's taxonomy [33]. Examples of YouTube problems are detailed in a number of publications [14, 24, 28]. Based on the characteristics of problems, YouTube problems can be grouped into various types (Table 1.1)

Table 1.1. Problem types, taxonomy verbs, and examples generated using YouTube videos.

Problem Type	Bloom's verbs	Example
Interpretation	Identify, Generalize, Classify, Translate	See Figure 1 – 1
Close-ended, Textbook-like	Demonstrate, Illustrate, Show, Solve	How it's made – Crayons (Figure 1 – 2)
Writing	Restate, Report, Summarize, Explain	Describe the 'event' in the video
Engineering Estimate (Fact or Fiction)	Compare, Analyze, Model, Relate	Big Water Slide detailed in [14].

Interpretation using diagrams play a key role in solving engineering problems. For example, interpretation problems can involve translating events in a video into a process flow diagram. Figure 1-1 is an example of an engineering diagram that shows the production of whiskey from the delivery of corn to the final product.

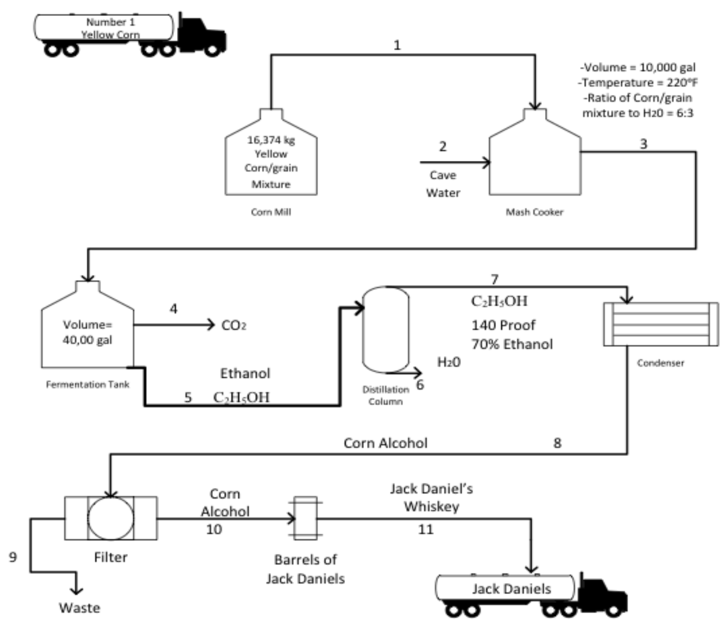


Figure 1-1: Example of student made process flow diagram from a material and energy balances course.

Close-ended problems have a single correct answer, use the video to identify a problem to reverse engineer, and incorporate physical values from the video in the problem. An example of a close-ended problems is titled "How it's made - crayons" (Figure 1-2). The type and scope of close-ended problems have varied greatly, from mimicking textbook problems, problems with single questions, problems with multiple parts, and sets of conceptual questions.

Video title: How it's made - Crayons
Video link: http://www.youtube.com/watch?v=m5f7NuGkhX0

Problem Statement:
(60 Points)

As stated in the video a large scale factory can produce 30,000 crayons/hr. This problem focuses on a much smaller scale factory, "Liberatore's Colors." The feed to the reactor is 150 mol/hr. Within this stream, there is 60 mole percent steric acid ($C_{18}H_{36}O_2$), 33.75 mole percent paraffin wax ($C_{20}H_{52}$), and the balance is Dr. Liberatore's own secret ingredient, the catalyst, Liberatorium. The reaction proceeds as follows,

$$110\ C_{18}H_{36}O_2 + 49\ C_{20}H_{52} \rightarrow 74\ C_{40}H_{82} + 220\ H_2O$$

Dr. Liberatore's sixth or even seventh sense can just tell that the single pass conversion of steric acid is 72%. After the crayons are made, the excess reactants continue to a separator where water is completely removed from the system with a small amount of Liberatorium. The composition of this waste stream is 99.8 mole percent water, and the balance Liberatorium. The crayons leave the separator as product. The fresh feed to the system is combined with a recycle stream that leaves the separator and contains the excess steric acid, paraffin wax, and Liberatorium; which then gets fed to the reactor. The fresh feed contains steric acid, paraffin wax, and Liberatorium.

- (5 points) Label the PFD with the component molar flow rates of each stream.
- (37 points) Find the flow rate of each component in the reactor effluent
- (12 points) Find the flow rate of the fresh feed
- (6 points) Find the volumetric flow rate of the Product Stream in SCMM.

Figure 1-2: Student written reaction-recycle problem for material and energy balances course [32].

A How It's Made video for crayons inspired an interesting homework problem for a reacting system with recycle (Figure 1-2). The problem statement is similar in length to the average textbook problem and includes a balanced chemical reaction, multiple parts/questions, and a process flow diagram, which was omitted from the Figure. The idealized reaction and separation scheme are common for sophomore level problems.

The first part of the problem statement for 'How it is made - crayons' (Figure 1-2) 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 in the course. The final part of the crayon problem involves finding the volumetric flow rate of the reactor effluent, fresh feed, and product stream in standard units (Figure 1-2). Since the product is a solid (crayons), use of standard temperature and pressure to compare volumetric flow rate reveals a misconception. Instructor can decide how to proceed, e.g., a timely class discussion on appropriate use of standard conditions and their effect (or lack of) on the density of gases, liquids, and solid.

More recently, YouTube problems are a set of new homework-style problems formed by reverse engineering a video to apply course concepts and utilized in subsequent classes or future offerings of a course. The current study considers YouTube problems as an alternative to textbook problems. The central hypothesis is that student-generated problems based on YouTube videos (YouTube problems) promote better problem-solving skills than traditional textbook problems. The goal of this research is to evaluate the efficacy of YouTube problems and examine the learning gains when students solve YouTube problems. Through the following research questions, this study seeks to examine the quality of problem-solving skills when solving YouTube problems:

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?
3. How do learning attitudes of students evolve over time?

To test the hypothesis examined by the research, this study requires multiple raters in assessment of problem solving and utilizes established survey tools to measure rigor of problems and learning attitudes of participants. Setup of the experimental intervention will be introduced first, each chapter follows a description of measurement tool and analysis of results and finally cross correlation of measurement from tools and conclusions.

1.3 Assessing Problem Solving

Measuring problem solving skill is a common practice among researchers. Many studies seeking to examine students' problem solving have often relied on grades given by classroom instructors [34]. Indicators, such as percentage of problems answered correctly or time spent solving a problem, have been used to represent problem solving ability [35]. Using grades as an indicator for problem solving proficiency is attractive due to its ease in collecting data with large numbers of students. However, using grades as a representation of problem-solving skill is often criticized because instructors consider many factors when grading that are not necessarily indicative of expert-like problem solving, such as fairness, motivating students, and the correctness of the end results [36, 37].

To avoid bias from utilizing grade as a measure of problem solving, some researchers have utilized self-reported measures [38-41]. Problem Solving Inventory (PSI) is an example of a self-reported survey that is common among the research community and has been applied in various fields [38, 42-47]. PSI is based on the underlying hypothesis that an individual's assessment of one's problem-solving competence influences a person's outcome in solving a problem and the whole process through problem solving [39, 48-50]. Challenges encountered with the PSI involves interpreting scoring scale due to indistinctness concerning what comprises "high" or "low" scores [44].

In search for better measurement for learning, Shute discovered the stealth assessment that involved the administering of tests without students knowing [51]. Stealth assessment, embedded in a game are virtually invisible to overcome test anxiety issues when taking tests and has been used to measure middle-school students' problem-solving skill [52, 53]. The process of developing the assessment begins by building a problem-solving competency model and identifying in-game indicators that would provide evidence about students' levels on various problem-solving facets. Indicators may include tracking in real time how long a player spends on a difficult problem or how often a player is willing to retry a problem that was previously failed [51-54].

Another strategy to measure problem solving is performance rubrics [37, 55]. Performance rubrics go beyond traditional assessments, i.e., only based on the correctness of students' answer. Rubrics, thus, provide instructors with valid and reliable information to monitor and offer feedback on students' progress related to specific criteria [56]. An example of a performance rubric is the PROCESS (Problem definition, Representing the problem, Organizing information, Calculations, Evaluation, Solution presentation and Self-assessment). PROCESS was developed by Benson and collaborators to measure the conceptual skill and analytical skill required in the process of problem solving in an engineering class [57-61]. PROCESS was designed to track every process 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 [57-61], the PROCESS rubric was adapted to assess problem solving in the current study.

1.4 Assessing Problem Difficulty

One of the most common practices of assessing problem difficulty is by making judgments based on an instructor's experience or intuition. Some factors that contribute to problem difficulty include: syntax variables, content and context variables, structure and complexity variables, and heuristic processes [62]. Judgement based on experience or intuition provides a qualitative measure. Despite knowledge of contributing factors to problem difficulty, intuition, or experience fails to provide a quantifiable measure problem difficulty [63].

Another technique for assessing problem difficulty is estimating the probability of successful problem solutions based on student performance, sometimes referred to as item analysis. Item analysis captures problem difficulty and the differentiation power of the answer option. Differentiation power is the capacity of a problem to differentiate between high and low performing students. Item difficulty is measured by the percentage of students who answered a question correctly while item discriminating power is the difference between the number of high-scoring students versus the number of low-scoring students who chose the same answer option [64]. Variables, such as time spent solving a problem or the number of attempts, can also indicate problem difficulty, but collecting data of this nature requires monitoring problem solving processes in real time [65]. However, assessing problem difficulty based solely on performance measures may not be the best means of evaluating problem difficulty since it does not consider how difficult the process is nor if there were extraneous factors that influenced success rates such as participants lacking pertinent prior information necessary to solve the problem [63].

A third measure of problem difficulty is based on students' perception. The three most 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) [66, 67]. Kevin et al, 1998 examined SWAT and the Modified Cooper-Harper Scale and found that the Modified Cooper-Harper Scale appeared to be less sensitive to changes in task difficulty than the SWAT [68]. NASA-TLX and SWAT possess similar features; for example, both have generic subscales, can be employed in any sphere, and do not interfere with task performance when administered after the task. SWAT has been criticized for being insensitive to low mental workloads.

Current study adopted NASA-TLX because of its ease in administration. In addition, NASA-TLX has been implemented as a measuring tool to assess problem difficulty in an Engineering class similar to present study and reported to have high validity.

Chapter 2

Materials and Methods

Research questions for this thesis focus on examining the effects of student-generated YouTube problems on the development of problem-solving skills in other students. The central hypothesis is that student-generated YouTube problems promote better problem-solving skills than traditional textbook problems. Therefore, research questions seek to examine the rigor of YouTube problems and efficacy of YouTube problems by investigating the effects that solving YouTube problems have on students' problem-solving ability and attitudes towards learning.

To answer these research questions, we employed a quasi-experimental, treatment/control group design (Table 2.1). 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 studying the same course content and using the same textbook, but not employing YouTube problems, served as a control group, thereby providing a quasi-experimental comparison design (Table 2.1).

Table 2.1**.** Summary of problem assignment to treatment and control group (first cohort).

Cohorts	**YouTube Problems**	**Textbook Problems**	**Created YouTube Problems**	**Class size**	**Randomly selected Students**
Treatment	9	10	1	90	30
Control	0	10	0	23	20

First year of intervention (2017-2018 academic year) constituted of cohort size of about 90 students in the treatment group from a large public university and 23 students in the control group at a private university. Control group learned the course material over a two-semester period unlike the treatment group that covered material balance and energy balance course over one semester. In order to balance sample sizes and make problem-solving assessment easier, we randomly selected 30 students' work from the treatment group to be scored.

Table 2.2**.** Number and type of homework problems assigned within certain 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

Faculty teaching a Material and Energy Balances course at both institutions had previous experience teaching the course, and they collaborated to ensure similar content delivery and the same control homework problems. Both treatment and control groups were assigned other homework questions from textbooks or other sources in addition to the homework problems as part of the study. Homework problems assigned to students covered a range of topics in Material and Energy Balance course (Table 2.2).

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, instructor 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 2.2).

For participants in the treatment group, homework assignments, normally 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 another measure of group equivalency. Since both treatment and control group use the same textbook [69], common problems were assigned. Students' homework solutions were utilized as scanned documents of hand-written solutions and scored. During periods where both groups completed textbook-only assignments, we compared the effects of solving YouTube problems on students' problem solving over time.

2.1 Assessing Problem Solving with 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 (Table 2.3). PROCESS evaluates both problem-solving process and final solution(s). PROCESS was modified to assess the problem-solving process for solved handwritten homework problems, which differs from its original use where participants solutions were collected on tablets and custom software to see erasing and other details [59, 61]. The tool was modified to suit material and energy balance problems. Each item in the revised PROCESS consists of four scaling levels ranging from 0 to 3 with zero being the minimum attainable score (Table 2.3).

Prior to scoring with the modified PROCESS, any identification regarding student or group identity was 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 do not reflect or have an effect on students' course grades.

Multiple raters used the PROCESS tool to assess problem solving. To ensure that raters utilize the PROCESS rubric in a consistent manner, raters initially assessed problem solving with PROCESS for the same subset of student responses. Raters' assessments were analyzed to determine how consistently raters measured problem solving ability. Traditional statistical (Cohen's kappa) and item response measures (Rasch many facets model) of inter-rater reliability were computed for five raters. The rater reliability assessment was carried out in two rounds until raters' scores exhibited consistency and accuracy in assessing student problem solving ability. This reliability assessment enhanced the validity of using the PROCESS tool to score MEB problems.

To determine whether the difference in group problem solving was statistically significant, we used the independent t-test as a measure of equivalency. Small p-values represents groups where differences are unlikely due to chance while large p-values denotes high probability that difference between groups is by chance. P-values smaller than 0.05 were considered as statistically significant.

Table 2.3. Modified process rubric for problem solving using handwritten solutions.

Problem Solving Process/ Category	Explicit Tasks Performed	Level of Completion				
		Missing 0 points	Inadequate 1 points	Adequate 2 points	Accurate 3 points	Error
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 ommisions	Fully documented execution tasks (Work showed evidence of relevant tasks)	Calculation error from extra equation
	coverted 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 multiparts problem answer	Provides mostly correct answers and units to all parts of the multiparts problem	Provides correct answers and units to all parts of the multiparts problem	presents wrong answers
	Correct units for answers to all parts of the problem.					wrong units
TOTAL SCORE						

2.2 Scoring using PROCESS Rubric

A problem titled 'Methanol reactor' is a typical textbook problem for a material and energy balance course (Figure 2-3). Here, the methanol reactor problem is categorized as a material balance with reaction. The first section of the problem requires students' ability to translate an engineering problem to a diagram while the second part of the problem tests knowledge in mass balance calculations, such as flowrates calculations, mole fractions, extent of reaction, and reaction conversions. The final part of the problem is a multiple-choice question that seeks to evaluate students' conceptual knowledge.

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

Figure 2-3: Problem statement a typical material and energy balance problem assigned to students as homework task.

Skills exhibited to earn 3 points on the P-section include identifying all unknown parameters, which include extent of reaction, molar flow rates of CO, N_2, H_2 and CH_3OH in the outlet stream (Figure 2-4a). However, a solution that identifies most but not all missing flowrates earns 2 points (Figure 2-4b). One point is assigned to a solution that identified few correct unknowns or mostly incorrectly unknowns (Figure 2-4c).

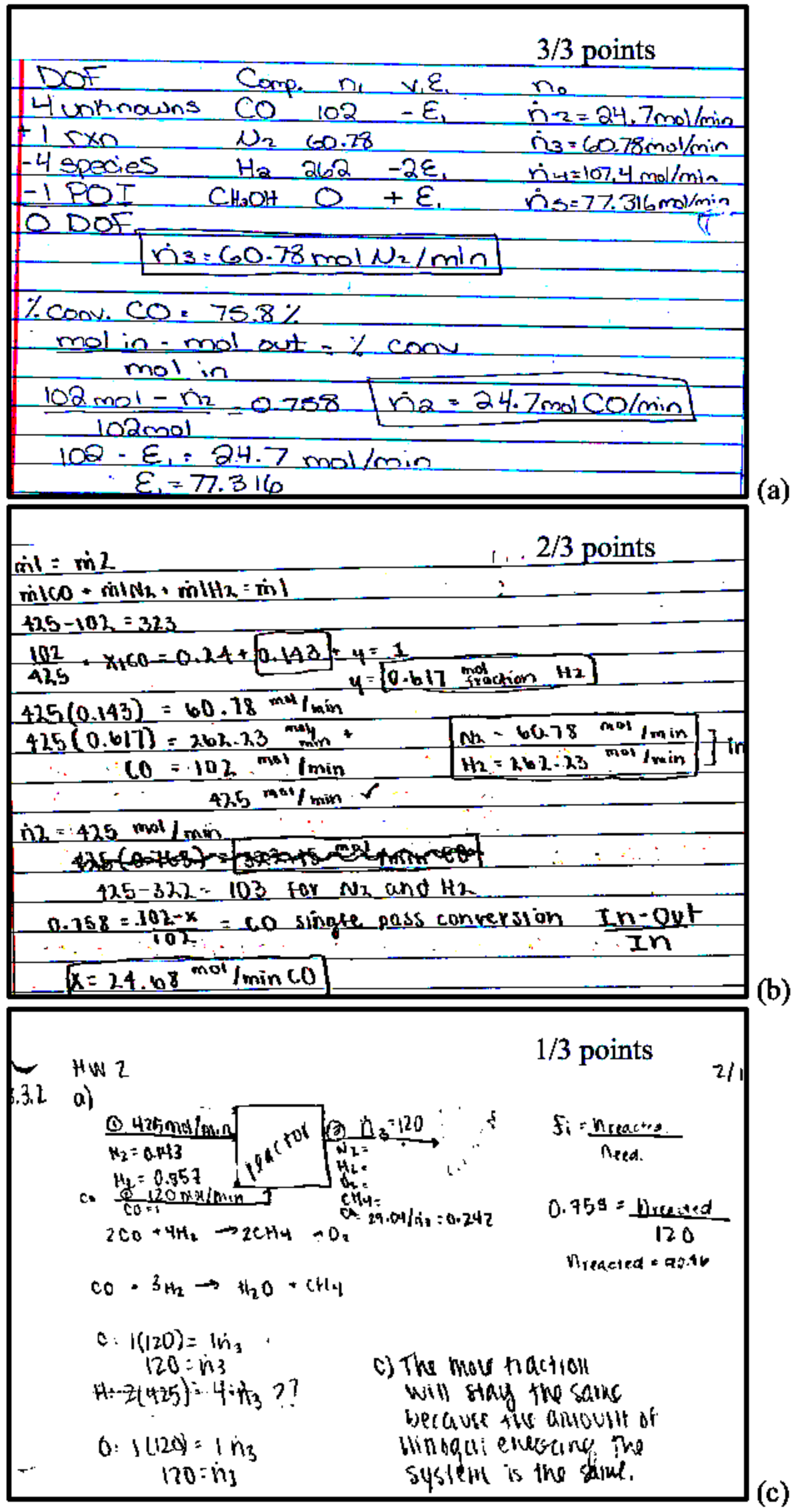


Figure 2-4: Samples of student written solutions to illustrate how problem identification stage (P) of PROCESS rubric is applied.

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 (Table 2.3). 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 (Figure 2-5a). When student's work contained most, but not all components in the streams, the solution earned 2 points (Figure 2-5b). 1 point is assigned when an attempt was made to draw the process diagram with few details (Figure 2-5c).

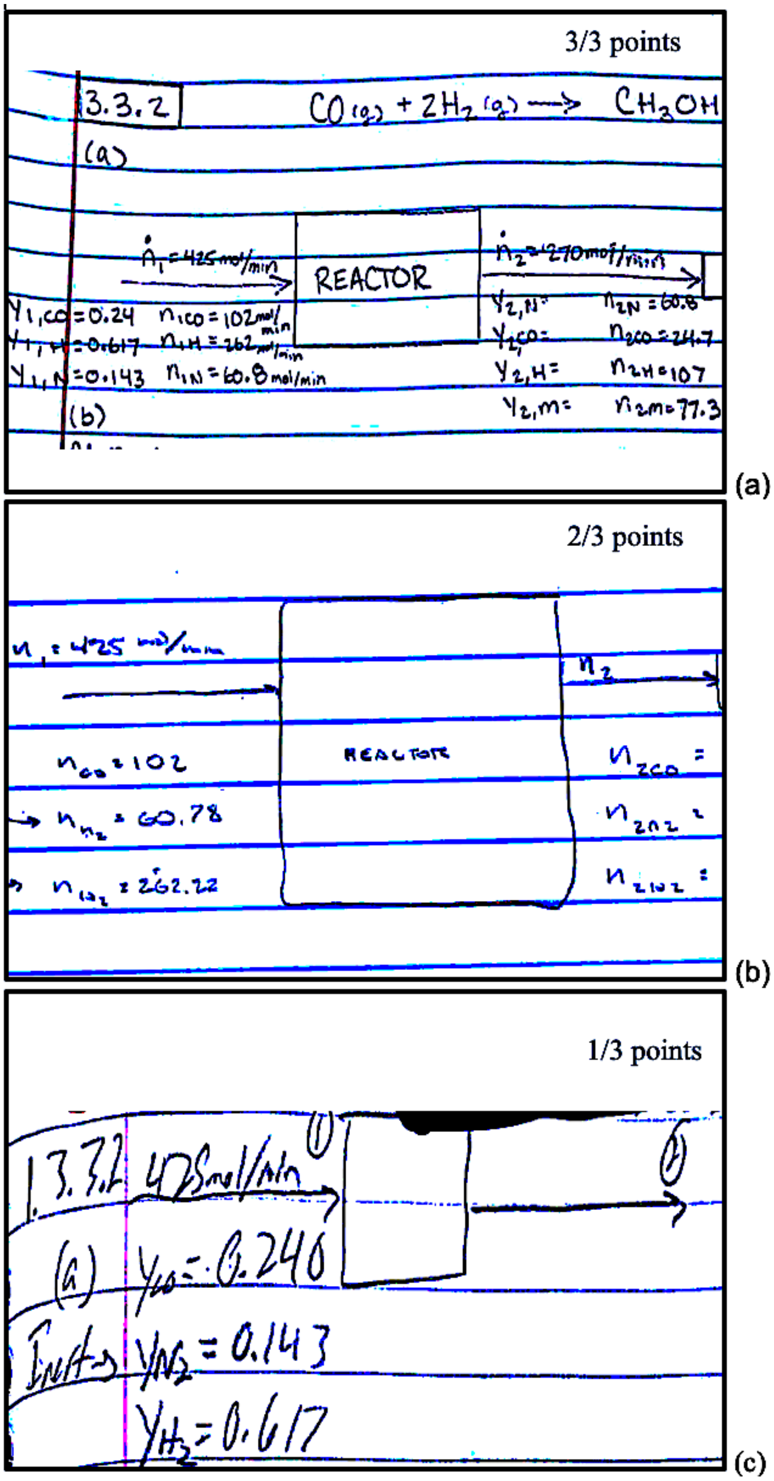


Figure 2-5: Samples of student written solutions to illustrate how problem representation stage (R) of PROCESS rubric is applied.

A written solution that correctly expresses all the important equations and material balances required to solve a problem earns 3 points in the O section. For the methanol reactor problem, solutions were expected to highlight an extra equation to find mole fraction of CO, correct expression for conversion, and atomic balances for C, H, O, and N to earn 3 points (Figure 2-6). A solution that expresses most of the important equations and some incorrect balances or equations is scored 2 points. When student's work expresses few equations with incorrect and incomplete atomic balances, 1 point is assigned.

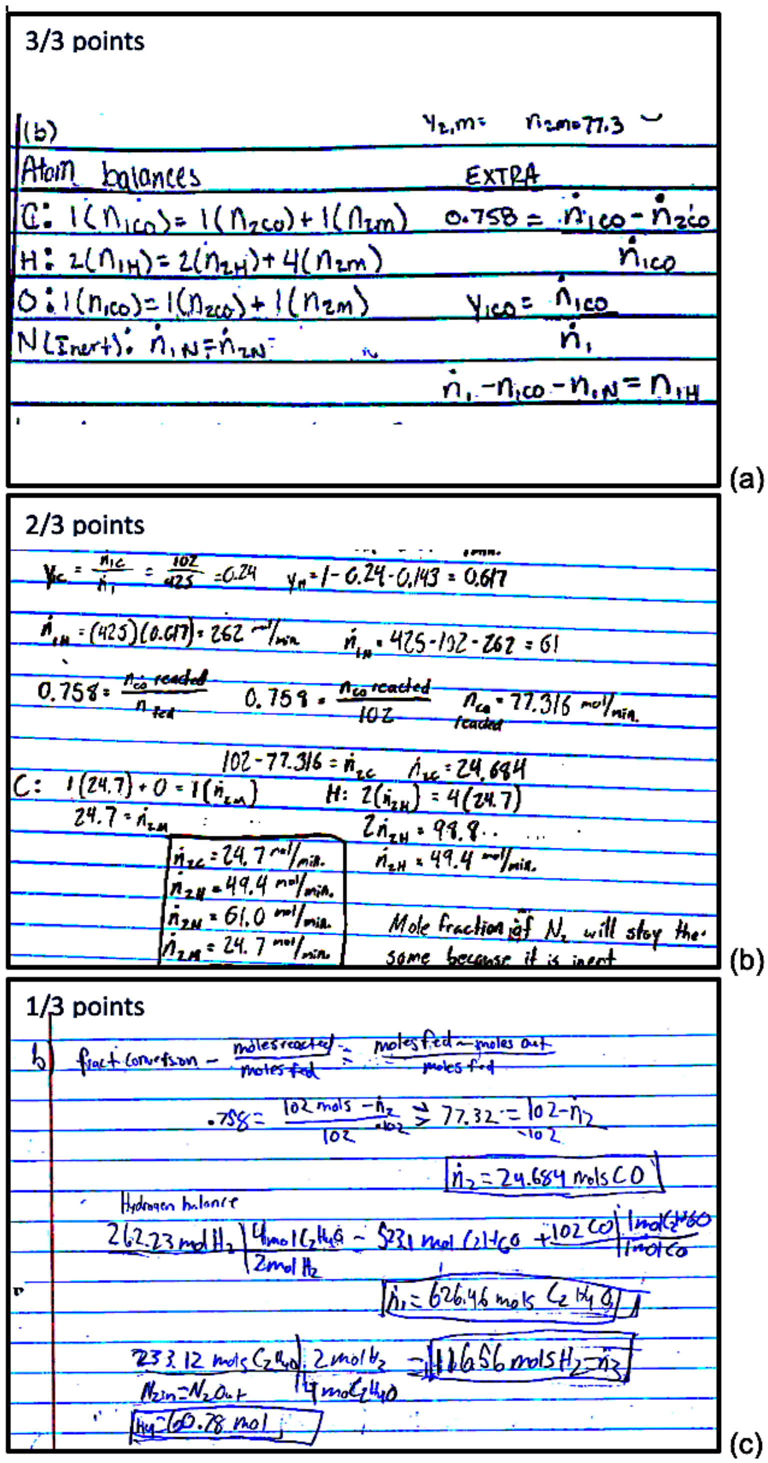


Figure 2-6: Samples of student written solutions to illustrate how organization stage (O) of PROCESS rubric is applied.

Achieving 3 points in the C section required detailed calculations leading to a solution for all parts of the problem (Figure 2-7a). A solution was scored 2 points when most of the calculations leading to the solution for all parts were completed (Figure 2-7b). In Figure 2-7b, solution failed to show how molar flow rate of H2 in the inlet stream was obtained. When a solution showed few calculations leading to the solution for multiple parts of the problem, solution was scored 1 point. For example, in Figure 2-7c, the solution failed to show work for obtaining molar flow rate for H2 and CH3OH in the outlet stream.

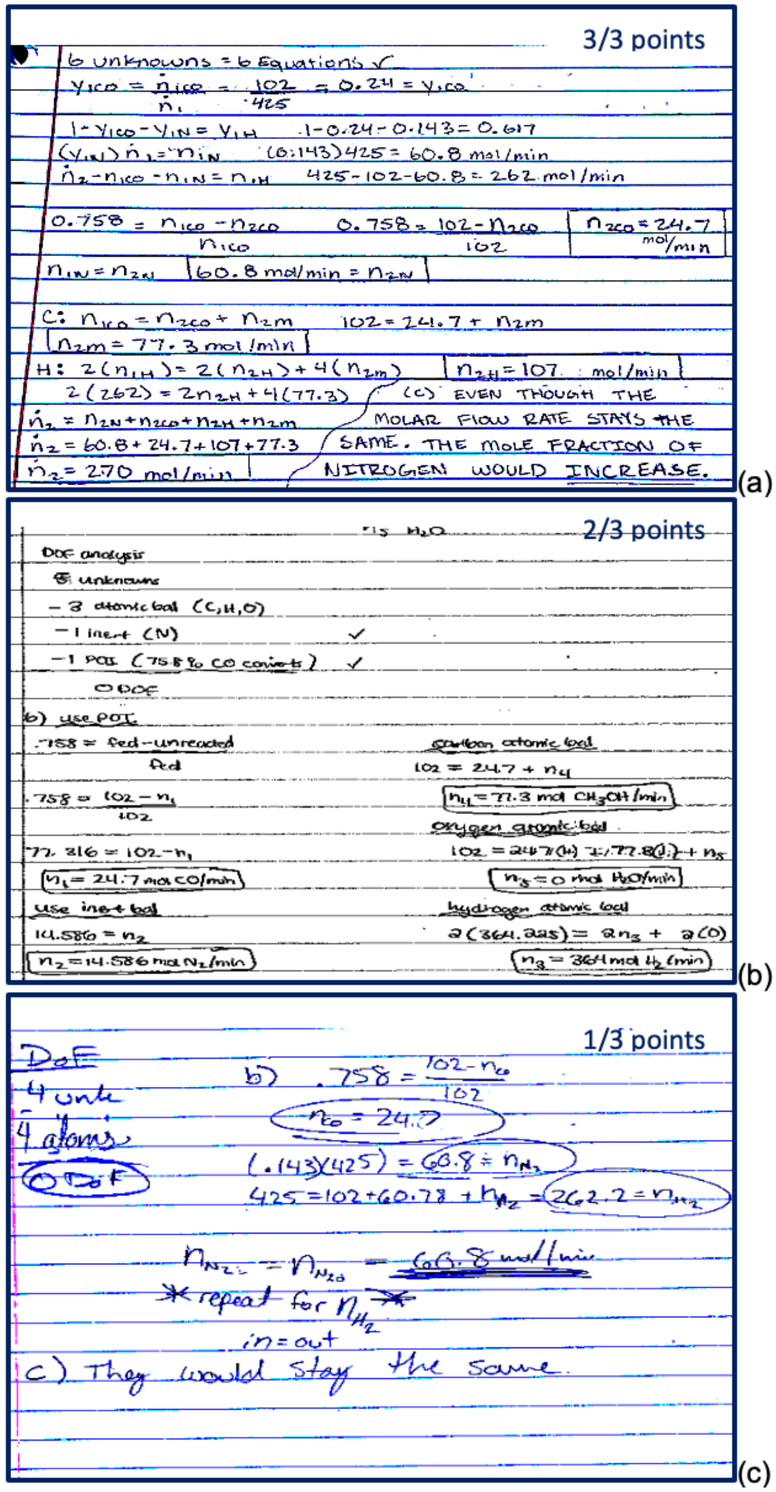


Figure 2-7: Samples of student written solutions to illustrate how calculation stage (C) of PROCESS rubric is applied.

Earning 3 points in solution completion required five answers, irrespective of the accuracy (Table 2.3). The answers required were exit flowrates of CO, H_2, N_2, CH_3OH and an answer to the multiple-choice question (Figure 2-6a). 2 points was earned when solution contained most of the answers. For example, Figure 2-6b provided three answers out of five, solution is scored 2 points in the completion stage. Figure 2-6c only showed answer to the molar flow rate of CO in the exit stream and did not answer the multiple-choice question, molar flow rate of N_2, CH_3OH, and H_2 in the outlet stream and was scored 1 point on solution completion.

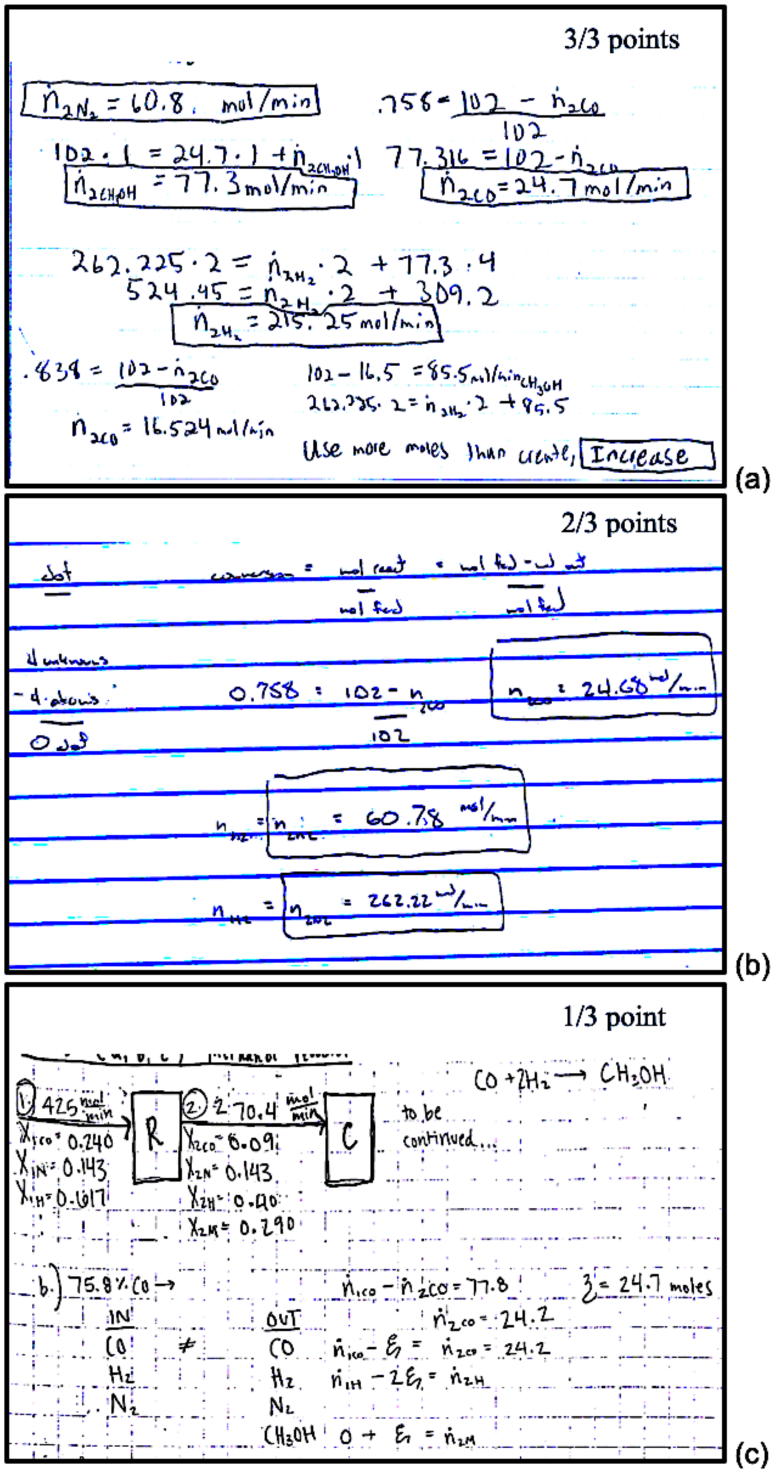


Figure 2-8: Samples of student written solutions to illustrate how solution completion stage (s) of process rubric is applied.

An accurate solution consisted of correct numerical values and units. When units were omitted or incorrect but with correct numerical values, answers were considered to be partially correct. When a student's work provided accurate answers with correct units to all parts of the problem statement, such work scored 3 points (Figure 2-9a). A solution with at mostly accurate answers was scored 2 points (Figure 2-9b). However, a solution was scored 1 point when it had few correct answers. Here, work did not show any answer for the molar flow rate of CH_3OH while the molar flow rate of H_2 in the exit stream and the answer to the multi-choice question were incorrect (Figure 2-9c).

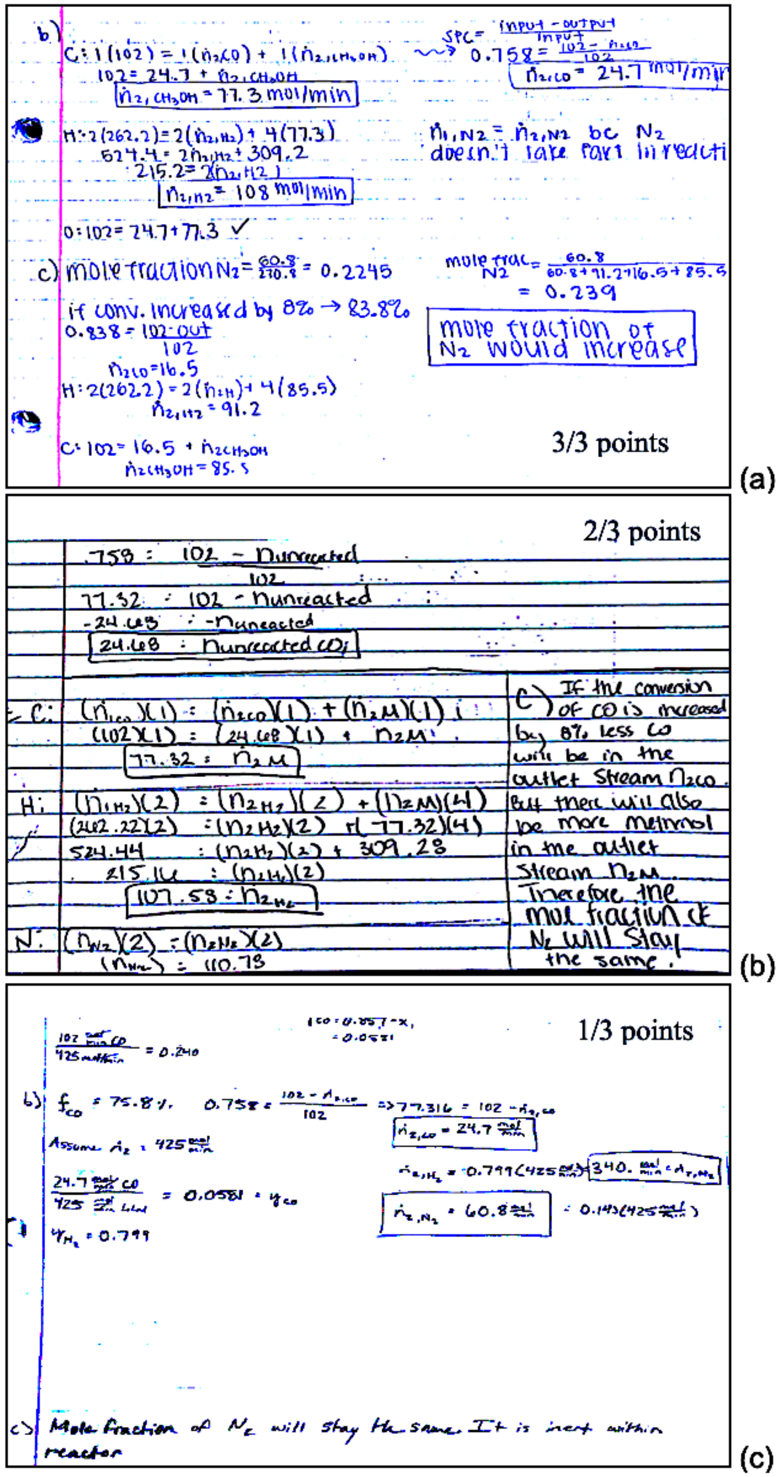


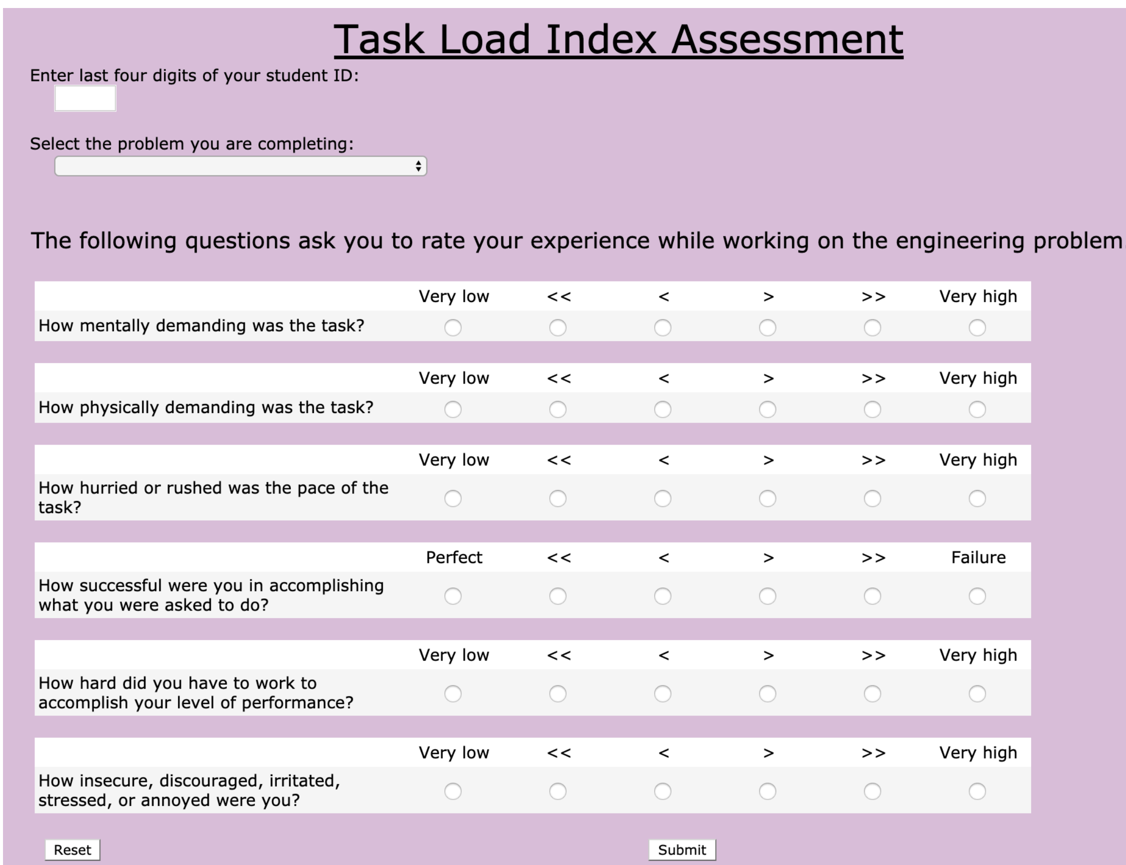
Figure 2-9: Samples of student written solutions to illustrate how accuracy stage (a) of process rubric is applied.

2.3 Assessing Problem Difficulty with 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. Sara et al. revealed that NASA TLX (Task Load Index) was an appropriate gauge of problem difficulty and yielded similar results to other accepted methods, such as the probability of success [63]. A key feature of the NASA TLX is the ease of implementation. NASA TLX is multidimensional and creates room for detailed analysis of workload source compared to other techniques. NASA TLX is also very portable and can be used in operational experiments [70]. NASA TLX was adopted to measure problem difficulty for this study because of its ease in administration [63].

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. The original NASA TLX measured workload in two stages. The first stage consists of participants ratings. Participants' ratings involve scoring each subscale of the NASA TLX in a continuous scale from least difficult and most demanding [63, 71-75]. The second stage of the TLX consist of a fifteen pairwise comparison of each subscale [73, 74].

Over the years, authors have projected and applied a modified version of the original scale by adding subscales, deleting some subscales, or redefining the existing subscales to improve the relevance to the target task or experimental questions [63]. Modifications on the NASA TLX often require establishing the validity, sensitivity, and reliability of the modified instrument before further surveys can be completed. The most common variation made to NASATLX excluded the weighting process that involved 15 pairwise comparison of the 6 constructs. Participants' rating in exclusion of the weight is often referred to as Raw TLX (RTLX) and has gained some popularity due to its ease in administration [76]. Another common adaptation is to analyze subscale ratings in place of a single overall workload score. The component ratings possess the capability of allowing the administrator identify the source of a workload or performance problem [73].



Task Load Index Assessment

Enter last four digits of your student ID:

Select the problem you are completing:

The following questions ask you to rate your experience while working on the engineering problem.

	Very low	<<	<	>	>>	Very high
How mentally demanding was the task?						
	Very low	<<	<	>	>>	Very high
How physically demanding was the task?						
	Very low	<<	<	>	>>	Very high
How hurried or rushed was the pace of the task?						
	Perfect	<<	<	>	>>	Failure
How successful were you in accomplishing what you were asked to do?						
	Very low	<<	<	>	>>	Very high
How hard did you have to work to accomplish your level of performance?						
	Very low	<<	<	>	>>	Very high
How insecure, discouraged, irritated, stressed, or annoyed were you?						

Reset

Submit

Figure 2-10: NASA TLX electronic survey used to collect workload data.

Current study utilized only the participants rating commonly referred to as RTLX to measure the rigor of problems. NASA TLX surveys were assigned alongside homework problems utilized for the study, i.e., not all homework problems for either cohort. NASA TLX was modified such that participants rating on subscales were on 6-point rating scale, where 1 is the least difficult and 6 being the most demanding (Figure 2-10). For each participant, overall workload was determined by summing scores for the 6 TLX questions resulting in an aggregate rating that ranges from 6 to 36. More demanding tasks earn higher scores. Difficulty of a problem was assessed by averaging participants TLX scores for the problem of interest. Analysis compared overall problem difficulty for different problem types and consistency in group responses.

2.4 Assessing Learning Attitudes with CLASS

To compliment PROCESS and NASA TLX, a measure of students' perception of problem-solving ability was employed. The Colorado Learning Attitudes about Science Survey (CLASS) [61, 77] was administered in a pre/mid/post participation design. Although CLASS was initially designed for physics, several studies have variations for chemistry, biology, astronomy and math [78].

This study utilized version 3 of CLASS for Physics, while the project team modified survey questions to match course content. The original version of CLASS was on a five-point Likert scale (strongly agree, agree, neutral, disagree to strongly disagree), but has been modified for the current project to a four-point Likert scale by eliminating the neutral option. Instructors tried to alleviate student's anxiety by informing students that correctness of the response will not affect their grades.

Table 2.4. Statements in the class version 3 [78].

1. A significant problem in learning engineering is being able to memorize all the information I need to know.
2. When I am solving an engineering problem, I try to decide what would be a reasonable value for the answer.
3. I think about the engineering I experience in everyday life.
4. It is useful for me to do lots and lots of problems when learning engineering.
5. After I study a topic in engineering and feel that I understand it, I have difficulty solving problems on the same topic.
6. Knowledge in engineering consists of many disconnected topics.
7. As engineers learn more, most engineering ideas we use today are likely to be proven wrong.
8. When I solve an engineering problem, I locate an equation that uses the variables given in the problem and plug in the values.
9. I find that reading the text in detail is a good way for me to learn engineering.
10. There is usually only one correct approach to solving an engineering problem.
11. I am not satisfied until I understand why something works the way it does.
12. I cannot learn engineering if the teacher does not explain things well in class.
13. I do not expect engineering equations to help my understanding of the ideas; they are just for doing calculations.
14. I study engineering to learn knowledge that will be useful in my life outside of school.
15. If I get stuck on an engineering problem my first try, I usually try to Figure out a different way that works.
16. Nearly everyone is capable of understanding engineering if they work at it.
17. Understanding engineering basically means being able to recall something you've read or been shown.
18. There could be two different correct values to an engineering problem if I use two different approaches.
19. To understand engineering, I discuss it with friends and other students.
20. I do not spend more than five minutes stuck on an engineering problem before giving up or seeking help from someone else.
21. If I don't remember a particular equation needed to solve a problem on an exam, there's nothing much I can do (legally) to come up with it.
22. If I want to apply a method used for solving one engineering problem to another problem, the problems must involve very similar situations.
23. In doing an engineering problem, if my calculation gives a result very different from what I'd expect, I'd trust the calculation rather than going back through the problem.
24. In engineering, it is important for me to make sense out of formulas before I can use them correctly.
25. I enjoy solving engineering problems.
26. In engineering, mathematical formulas express meaningful relationships among measurable quantities.
27. It is important for the government to approve new scientific ideas before they can be widely accepted.
28. Learning engineering changes my ideas about how the world works.
29. To learn engineering, I only need to memorize solutions to sample problems.

30. Reasoning skills used to understand engineering can be helpful to me in my everyday life.
31. We use this statement to discard the survey of people who are not reading the questions. Please select strongly agree-option 4 for this question to preserve your answers.
32. Spending a lot of time understanding where formulas come from is a waste of time.
33. I find carefully analyzing only a few problems in detail is a good way for me to learn engineering.
34. I can usually figure out a way to solve engineering problems.
35. The subject of engineering has little relation to what I experience in the real world.
36. There are times I solve an engineering problem more than one way to help my understanding.
37. To understand engineering, I sometimes think about my personal experiences and relate them to the topic being analyzed.
38. It is possible to explain engineering ideas without mathematical formulas.
39. When I solve an engineering problem, I explicitly think about which engineering ideas apply to the problem.
40. If I get stuck on an engineering problem, there is no chance I'll figure it out on my own.
41. It is possible for engineers to carefully perform the same experiment and get two very different results that are both correct.
42. When studying engineering, I relate the important information to what I already know rather than just memorizing it the way it is presented.

CLASS consisted of 42 survey questions (Table 2.4). Question 31 is an attention-check, and surveys with incorrect answers for question 31 were not used. CLASS usually required less than 10 minutes for respondents to complete. Questions from CLASS can be grouped into eight parts including: Personal Interest, Real World Connection, Problem Solving General, Problem Solving Confidence, Problem Solving Sophistication, Sensemaking/Effort, Conceptual Understanding, and Applied Conceptual Understanding (Table 2.5) [78]. Each category consisted of four to eight statements that describe a specific aspect of student thinking. These subgroupings of the CLASS included 27 of the 42 statements.

Table 2.5. Categories of survey questions from class [78].

Categories	Statements comprising	Number of statements
Real World Connection	28, 30, 35, 37	4
Personal Interest	3, 11, 14, 25, 28, 30	6
Sense Making/ Effort	11, 23, 24, 32, 36, 39, 42	7
Conceptual Connection	1, 5, 6, 13, 21, 32	6
Applied Conceptual Thinking	1, 5, 6, 8, 21, 22, 40	7
Problem Solving General	13, 15, 16, 25, 26, 34, 40, 42	8
Problem Solving Confidence	15, 16, 34, 40	4
Problem Solving Sophistication	5, 21, 22, 25, 34, 40	6
Not Scored	4, 7, 9, 31, 33, 41	6

For the purpose of analysis, strongly agree and agree were considered to be the same selection and similarly, disagree and strongly disagree were considered to be the same selection (Table 2.6) [78]. The CLASS survey was scored for overall learning attitude and for its eight categories. The overall score included 36 statements out of the total 42 statements in the CLASS. The remaining six statements were five questions that had no consensus expert response and an attention check question.

Table 2.6. Class expert response key [78].

Question	Desired Response	Question	Desired Response	Question	Desired Response
1	D	15	A	29	D
2	A	16	A	30	A
3	A	17	D	31	D only
4	N	18	D	32	D
5	D	19	A	33	N
6	D	20	D	34	A
7	N	21	D	35	D
8	D	22	D	36	A
9	N	23	D	37	A
10	D	24	A	38	A
11	A	25	A	39	A
12	D	26	A	40	D
13	D	27	D	41	N
14	A	28	A	42	A

A = experts agree or strongly agree.

D = experts disagree or strongly disagree.

N = there is no consensus expert response.

D only = "disagree" is the only correct answer to this question

CLASS score was obtained by determining, for each student, the percentage of responses for which the student agrees with the experts' response. For example, statement 1 in CLASS survey states "A significant problem in learning engineering is being able to memorize all the information I need to know." The desired response for question 1 was either disagree or strongly disagree. If a participant response agreed or strongly agreed with statement 1, participant gets 0 points on statement 1. However, if a participant disagrees or strongly disagree with statement 1, participant would earn 1 point on statement 1. CLASS scores for each participant represented number of times a participant gave a desired response expressed as a percentage. The procedure of calculating CLASS scores was used considering the relevant statements for each category (Table 2.5). Analysis compared group CLASS scores of students at the beginning and end of course. Normalized gain and Cohen d effect size analysis was used to determine magnitude of change in learning attitudes of participants for pre and post-test.

Chapter 3

Results and Discussion

We will start by analyzing findings from NASA TLX survey collected for students in two cohorts by comparing NASA TLX scores for two homework types considered in this study and identifying significant contributors to perception of problem difficulty. Analysis of NASA TLX survey will be followed by a detailed comparison of PROCESS scores for students in first year cohort. Stages in PROCESS will be examined and correlation between NASA TLX scores and PROCESS scores will be examined for homework problems. CLASS scores collected for two cohorts will be examined for possible changes across period when course was administered.

Table 3.1. Summary of NASA TLX survey responses for 10 textbook problems completed by participants in control group over two years cohort.

	Mean		Number of participants		
Textbook problems	Year 1	Year 2	Year 1	Year 2	p
TB1	18.1± 6.2	18.1 ± 6.1	24	27	1.0
TB2	21.1 ± 5.0	18.4 ± 5.7	20	22	0.1
TB3	25.6 ± 4.5	22.4 ± 6.0	13	26	0.1
TB4	19.6 ± 5.1	21.9 ± 7.5	22	20	0.2
TB5	23.4 ± 4.6	22.3 ± 5.6	21	20	0.5
TB6	24.7 ± 4.9	22.2 ± 5.2	16	21	0.1
TB7	23.4 ± 7.0	18.8 ± 4.5	15	26	0.01*
TB8	18.5 ± 6.4	15.3 ± 6.0	18	24	0.1
TB9	23.5 ± 4.5	19.9 ± 5.8	11	18	0.09
TB10	27.9 ± 4.5	26.1 ± 5.1	8	13	0.4

*Indicates a statistically significant difference.

Table 3.2. Summary of NASA TLX survey responses for 10 textbook problems completed by participants in treatment group over two years cohort.

	Mean		Number of participants		
Textbook problems	Year 1	Year 2	Year 1	Year 2	p
TB1	17.5 ± 5.4	17.3 ± 5.9	86	88	0.8
TB2	19.2 ± 6.4	19.3 ± 5.5	84	84	0.9
TB3	23.6 ± 6.2	23.5 ± 6.1	72	72	0.9
TB4	18.7 ± 5.9	17.2 ± 5.8	82	68	0.1
TB5	22.8 ± 5.5	21.8 ± 5.4	58	59	0.3
TB6	24.1 ± 5.1	24.4 ± 6.0	29	58	0.8
TB7	19.9 ± 5.4	19.0 ± 5.3	50	56	0.4
TB8	13.8 ± 6.5	13.7 ± 5.7	47	62	0.9
TB9	20.6 ± 5.8	-	41	-	-
TB10	24.2 ± 6.0	25.2 ± 4.9	40	51	0.4

Table 3.3. Summary of NASA TLX survey responses for 9 YouTube problems completed by participants in treatment group over two years cohort.

	Mean		Number of participants		
YouTube problems	Year 1	Year 2	Year 1	Year 2	p
YT1	22.8±5.8	20.9±6.0	78	70	0.05
YT2	25.6±5.7	22.5±6.7	66	67	0.005*
YT3	22.4±6.2	19.4±6.8	75	72	0.006*
YT4	20.8±6.5	20.0±5.4	63	71	0.4
YT5	17.7±6.2	16.1±4.9	66	78	0.09
YT6	18.9±6.3	17.9±5.2	56	68	0.3
YT7	17.9±5.3	16.8±5.7	50	65	0.3
YT8	18.3±5.8	17.8±6.6	45	68	0.7
YT9	21.9±5.5	21.0±7.4	42	60	0.5

*Indicates a statistically significant difference.

Based on students' ratings from NASA-TLX, most participants in control group responded to survey consistently when completing each Textbook problem (Table 3.1). Significant difference in ratings was seen (Table 3.2) when students in the control group solved 1 textbook problem (TB7). When Treatment group solve Textbook problems, comparison of responses between first- and second-year cohorts yields large p-values indicating similarity of responses (Table 3.3). However, significant difference ($p = 0.005$) in ratings was reported for two YouTube problems (YT2 and YT3).

Table 3.4. Summary of NASA TLX survey responses for 10 textbook problems completed by participants in control and treatment group over two years cohort.

	Mean		Average number of participants		
Textbook problems	Control	Treatment	Control	Treatment	p
TB1	18.1 ± 6.1	17.4 ± 5.7	26	87	0.6
TB2	19.7 ± 5.3	19.3 ± 6.0	21	84	0.8
TB3	24.0 ± 5.3	23.6 ± 6.1	20	72	0.8
TB4	20.7 ± 6.3	18.0 ± 5.8	21	75	0.07
TB5	22.9 ± 5.1	22.3 ± 5.5	21	59	0.7
TB6	23.4 ± 5.1	24.2 ± 5.6	19	44	0.6
TB7	21.1 ± 5.8	19.4 ± 5.4	21	53	0.2
TB8	16.9 ± 6.2	13.8 ± 6.1	21	55	0.05
TB9	21.7 ± 5.1	20.6 ± 5.8	15	41	0.5
TB10	27.0 ± 4.8	24.7 ± 5.4	11	46	0.2

Since we have reasonable agreement between data collected for two years cohorts, participants ratings were aggregated. Based on combined ratings from year 1 and year 2 cohorts, we compared ratings across different groups to study how participants in Treatment and Control group utilize the NASA TLX survey. Here, we compare NASA TLX ratings for Textbook problems since these problems were commonly completed by both groups. Students ratings reveals that difficulty vary from problem to problem (Table 3.4). Treatment group rated every problem to be slightly more difficult compared to ratings from Control group. However, large p-values indicated that the difference in ratings between groups are not significant.

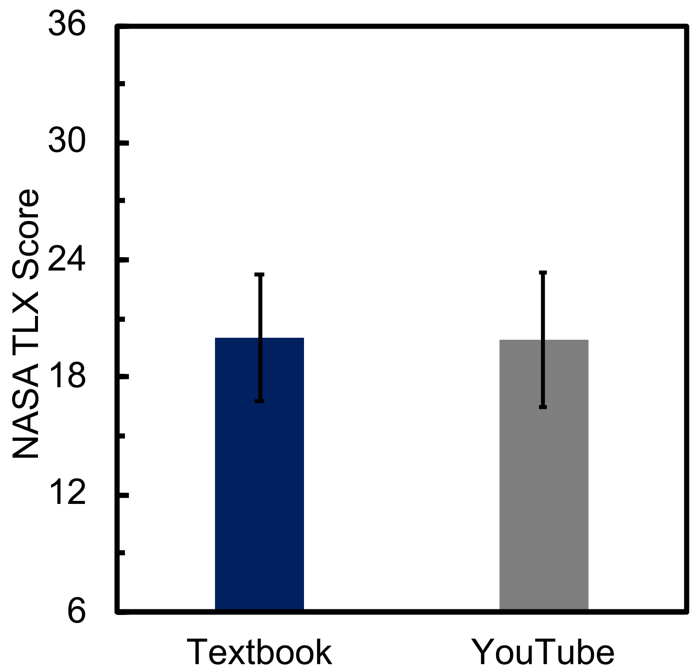


Figure 3-11: Students difficulty rating for 19 problems categorized under two problem types. Difficulty ratings is as a function of NASA TLX responses from ~200 students in Treatment group for two academic years.

Table 3.5. Overall perceived difficulty measured with NASA TLX when students solve YouTube and textbook problems.

	Textbook	YouTube
Mean	20.0±3.5	19.9±2.5
Number of problems	10	9
p	0.92	

Based on average of aggregated ratings on all problems within each problem type (Textbook and YouTube problem), Treatment group compared both problem types (Figure 3-11). Slightly larger standard deviations were reported for Textbook problems (3.5 vs 2.5) denoting wider variation of difficulty levels (Table 3.5). Participants in the Treatment group reported that differences in difficulty ratings of Textbook and YouTube problems were not significant as evidenced by $p = 0.92$.

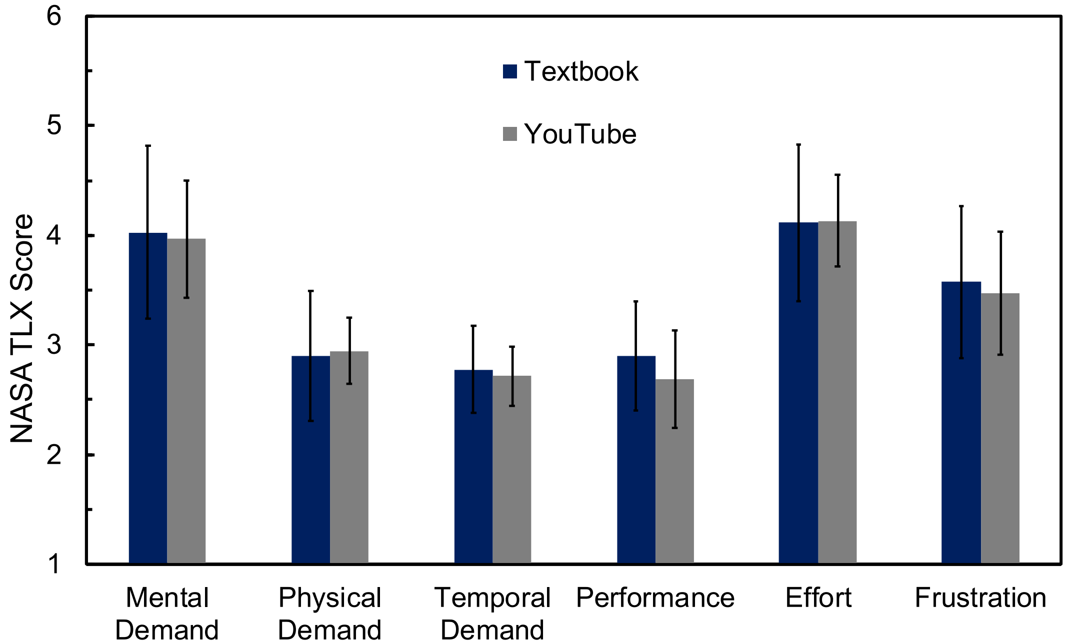


Figure 3-12: Students' perception on constructs of the NASA TLX when solving material balance problems from 10 textbook and 9 YouTube problems.

Table 3.6. Students' perception of difficulty on each component of NASA TLX when solving material balance problems from textbook and YouTube problem.

NASA TLX components	Mean		Number of problems		p	Rank
	Textbook	YouTube	Textbook	YouTube		
Mental demand	4.0±0.8	4.0±0.5	10	9	1.0	2
Physical demand	2.9±0.6	2.9±0.3	10	9	1.0	4
Temporal demand	2.8±0.4	2.7±0.3	10	9	0.6	6
Performance	2.9±0.5	2.7±0.4	10	9	0.4	5
Effort	4.1±0.7	4.1±0.4	10	9	1.0	1
Frustration	3.6±0.7	3.5±0.6	10	9	0.7	3

Item analysis of students rating of problems allows us to identify the impact of different items on problem difficulty. NASA TLX ratings for 19 problems from ~ 200 students in the treatment group collected for year 1 and year 2 cohorts, revealed that textbook problems and YouTube problems had similar rigor for every construct of the NASA TLX (Table 3.6). However, participants rated Textbook problems 4% more difficult than YouTube problem in terms of performance, temporal demand and frustration. Large p-values indicated that differences in ratings of problem types were not statistically significant (Table 3.6). Mental demand, contributed over 30% more to problem difficulty than other types demands (physical and temporal demands) measured by the NASA TLX (Figure 3-12). Students report that effort was the most pronounced contributor (10% more than stress due to frustration and performance) to stress when solving material and energy balance problems. Physical demand did not show high contribution to problem difficulty because solving homework problems did not require physical work besides either watching a video and/or writing a solution. Similarly, temporal demand was not a major contributor to the overall demands when solving homework problem owing to sufficient amount of time (one week) given to students to complete each homework set.

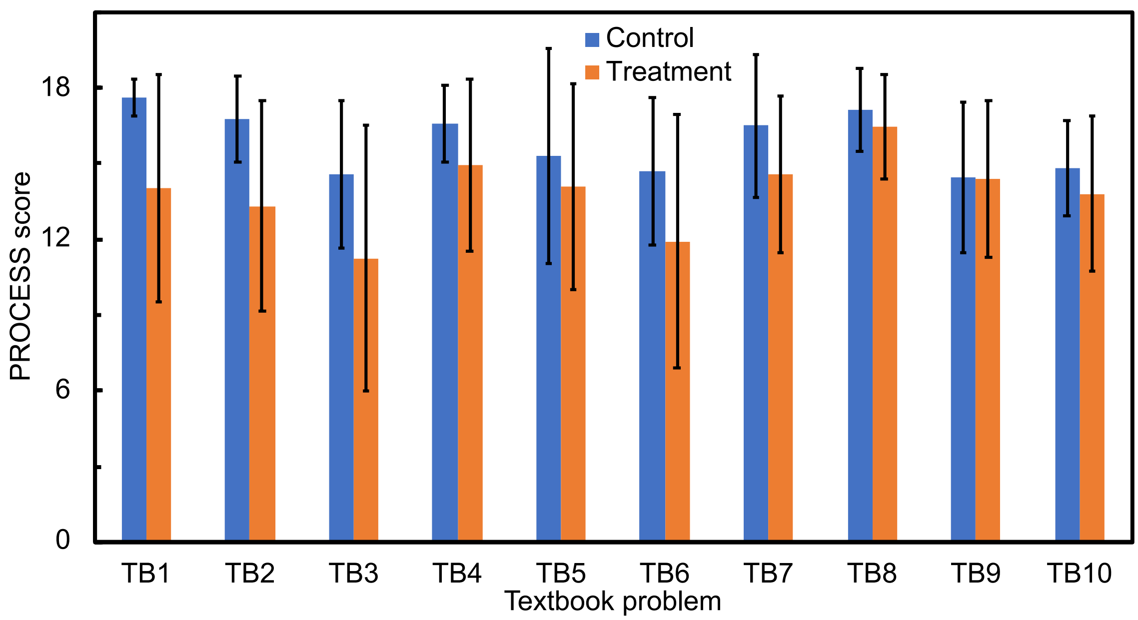


Figure 3-13: Overall problem-solving ability of participants in treatment and control group when solving 10 textbook problems.

Table 3.7. Summary of process assessment for textbook problems obtained at the beginning to the end of study.

	Mean		number of participants		
Textbook problems	Control	Treatment	control	treatment	p
TB1	17.6 ± 0.7	14.0 ± 4.5	19	11	0.002*
TB2	16.8 ± 1.7	13.3 ± 4.2	19	25	0.001*
TB3	14.6 ± 2.9	11.2 ± 5.3	24	27	0.008*
TB4	16.6 ± 1.5	14.9 ± 3.4	14	28	0.08
TB5	15.3 ± 4.3	14.1 ± 4.1	18	27	0.4
TB6	14.7 ± 2.9	11.9 ± 5.0	16	25	0.05
TB7	16.5 ± 2.8	14.6 ± 3.1	19	26	0.04*
TB8	17.1 ± 1.7	16.4 ± 2.1	17	29	0.3
TB9	14.4 ± 3.0	14.4 ± 3.1	12	43	1
TB10	14.8 ± 1.9	13.8 ± 3.1	9	24	0.4

*Indicates a statistically significant difference.

Having established difficulty of problems via NASA TLX, another parameter of high import was problem-solving skills as it could be influenced by type of practice problems. PROCESS scores provided a measure for problem solving ability. PROCESS scores for each student on a particular problem were obtained by summing the 6 category scores of PROCESS rubric. PROCESS scores can range from 0 to 18 (Figure 3-13). An overall PROCESS score was obtained for each of the ten Textbook problems by calculating the average performance of the treatment and control groups on each problem, respectively (Figure 3-13). Approximately 1500 PROCESS scores were collected for 50 participants solving Textbook problems (Table 3.7). Students' solutions are not assessed when students turn in solution for a different problem. At least three raters overlap when scoring each student solution with PROCESS. Raters spent 10 to 15 minutes scoring each participant's solution with PROCESS. Control group earned higher PROCESS scores than the treatment group in all textbook problems. Problem solving as indicated by PROCESS scores differs from problem to problem. Group mean PROCESS scores ranged between 11 to 18.

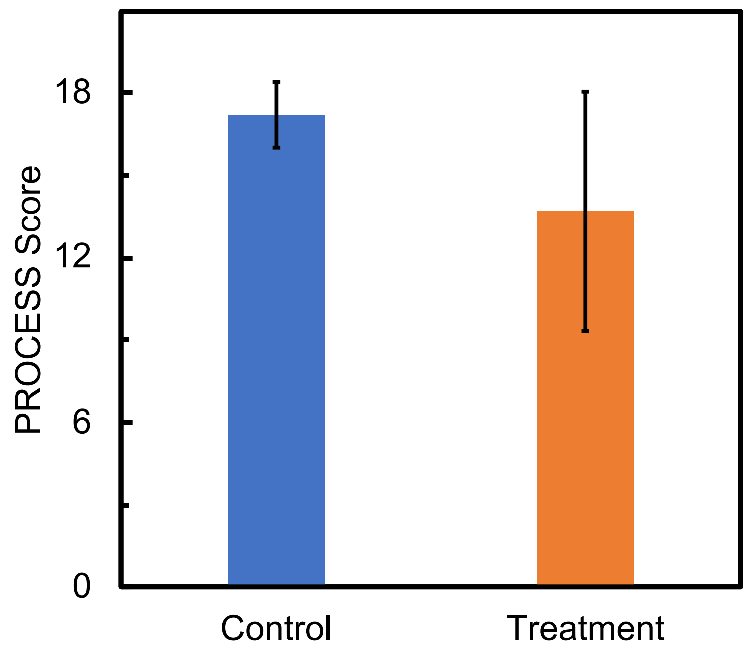


Figure 3-14: Average PROCESS scores by participant group at the beginning of the study for two problems: volume percent (TB1), and isopropanol mixing (TB2).

First, we investigated the equivalence of the treatment and control group based on their problem-solving ability when solving textbook problems at the beginning of the course. Treatment and control participant groups completed two textbook problems early in the term (Figure 3-4). PROCESS scores for problems completed earlier in the study revealed that control group earned higher PROCESS scores than the treatment group (Figure 3-4). The mean PROCESS score of the control group was 25% higher than the treatment group (17.2 vs 13.7). Furthermore, treatment group showed a larger standard deviation than the control group. Differences in the mean PROCESS scores were quantified by performing a two-tailed t test, which resulted in a p-value 0.002, which is statistically significant. Thus, problem solving abilities of control and treatment groups at the beginning of study were different with better solving ability in the control group. Most students in the treatment group were in their second semester of the freshman year compared to first semester of the sophomore year for the control group, which might be an advantage for the students in the control group.

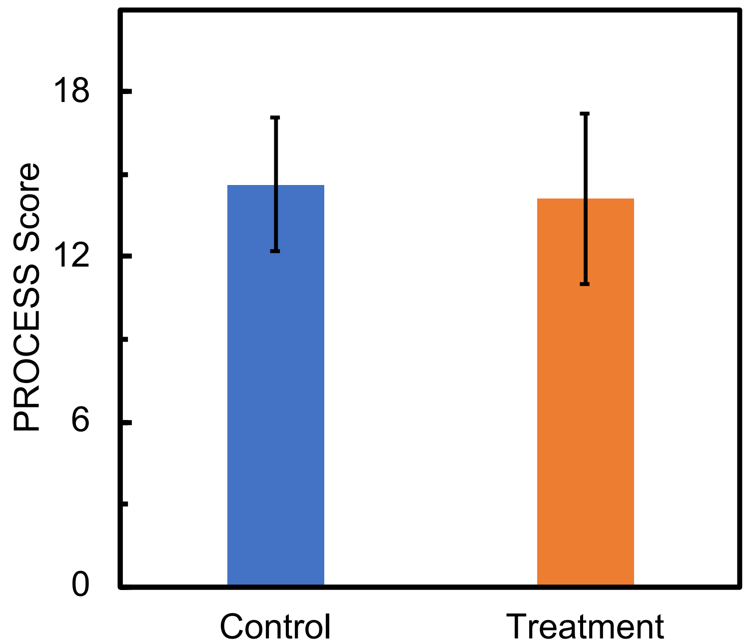


Figure 3-15: Average PROCESS scores versus participant group at the end of the study for two energy balance problems: (TB9) Ice ammonia and (TB10) Ethane combustion.

In contrast to observations in the beginning of course, students in the treatment group exhibited similar problem-solving ability to the control group near the end of the study (Figure 3-15). Mean PROCESS score of the control group was only 4% higher than the treatment group (14.6 vs 14.1). The treatment group also exhibited slightly wider range of scores with larger standard deviations than the control group (Table 3.7). Differences in mean PROCESS scores were not statistically significant ($p = 0.6$). Therefore, control and treatment groups for first year cohort demonstrated similar problem-solving ability at the end of the study.

Table 3.8. PROCESS scores earned by treatment and control group across semester. Mean PROCESS scores were calculated from average PROCESS scores earned in Textbook problems, 2 completed at the beginning (pre) and 2 at end (post) of course.

	Mean		Average number of participants		
	Control	Treatment	control	treatment	p
pre	17.2 ± 1.2	13.7 ± 4.4	19	18	0.002*
post	14.6 ± 2.4	14.1 ± 3.1	11	34	0.63
Normalized gain	-0.15	0.10			
Effect size (Hedge's g)	-1.5	0.11			

*Indicates a statistically significant difference.

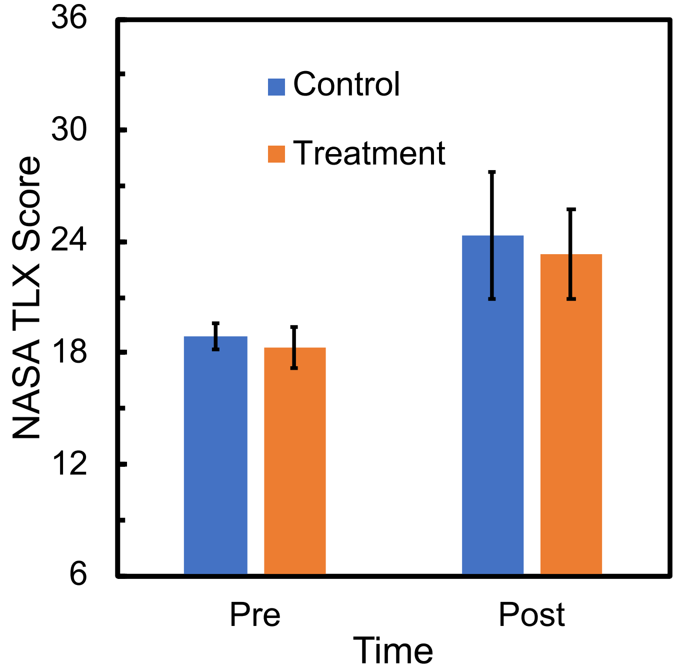


Figure 3-16: Overall NASA TLX scores reported by students in 2 years cohort for Textbook problems, 2 completed at the beginning (pre) and 2 at end (post) of course. Error bars indicate standard deviation of mean NASA TLX scores for problems considered.

Comparing average PROCESS scores in the beginning of the semester(pre) to those at the end (post) of the semester (Table 3.8), the average PROCESS scores for control group dropped by 15% near the end of the study (17.2 vs 14.6). One possible reason might be due to a 30% increase (18.6 vs 23.8) in the difficulty of the problems as reported through students' ratings (Figure 3-16). The first set of problems covered earlier in the course concepts of volume percent/mole ratio calculation and basic mass balances for non-reacting systems. On the other hand, later problems covered both material and energy balance concepts for reacting systems. Towards the end of course, a convergence of PROCESS scores between control and treatment groups evidenced by large p-values (0.63). Large p-values indicated that YouTube problems do not have a detrimental effect on students' problem-solving skills. Also, 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 (Table 3.8). Additional scoring and a second annual cohort were collected to more clearly answer these questions in future work.

An examination of within group changes provides another perspective on the effects of the YouTube problems on problem-solving ability (Table 3.8). One way to quantify changes is through normalized gain in average PROCESS score. Students in the treatment group showed a positive normalized gain (0.1) in problem solving ability unlike the control group that showed no gain (-0.15). Hedges' g effect size provides another measure to quantify changes within group by incorporating the mean PROCESS scores and pooled standard deviations. Effect size values ranges from small (<0.3), medium (0.3 to 0.6), to large (>0.7). The control group exhibited a negative, large effect size change in problem-solving skill (Hedges' g = -1.5 whereas the treatment group on the other hand, realized a small positive effect size gain (Hedges' g = 0.1) between the initial problems and the later problems. Students who completed YouTube problems showed small gains in their problem-solving skill level throughout the semester. The control group, however, once faced with more rigorous problems (Figure 3-16), displayed decreased level of problem-solving (Table 3.8). Results suggests that the YouTube problems could be contributing to long term, positive effects on student problem-solving ability.

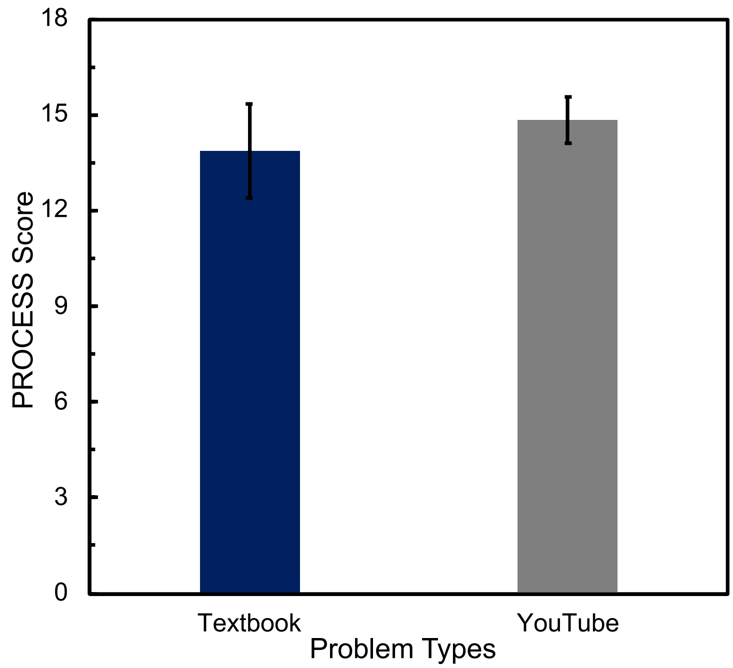


Figure 3-17: Overall problem-solving ability when students solve 9 YouTube and 10 textbook problems.

Students in the treatment group displayed slightly higher PROCESS scores when solving YouTube problems than when they solve textbook problems (Figure 3-17). Students showcased 7% better problem-solving abilities for YouTube problems than when completing Textbook problems (mean of 14.8 vs 13.9). The difference is not statistically significant (p = 0.12) (Figure 3-17). Higher scores with YouTube problems might be due to the video component which is integrated into YouTube problems. Visual effect achieved with video could be responsible for helping students understand problems better and in turn display higher problem-solving capability. Overall, students exhibited a similar problem-solving skill for YouTube and Textbook problems.

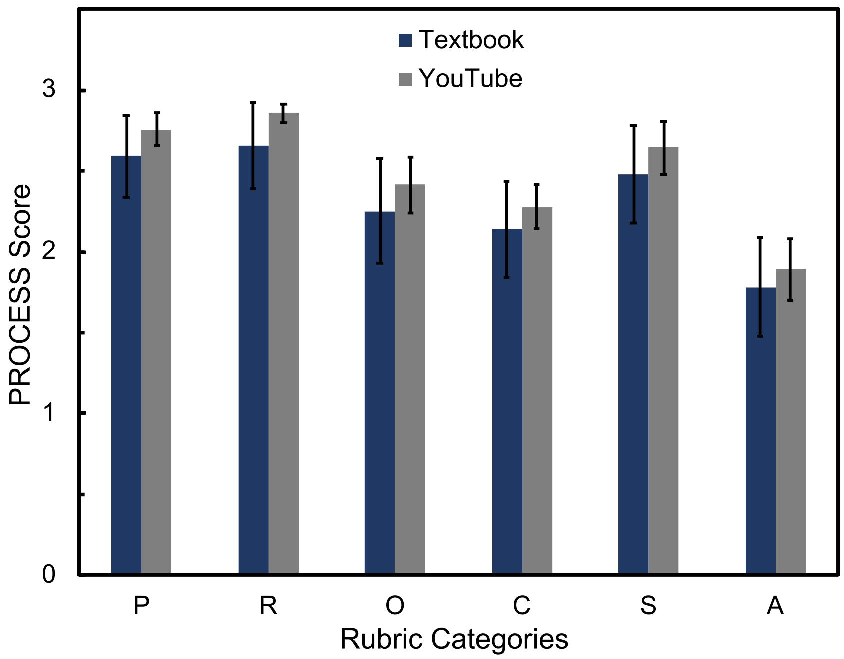


Figure 3-18: Problem solving ability of the treatment group when solving 10 textbook and 9 YouTube Problems.

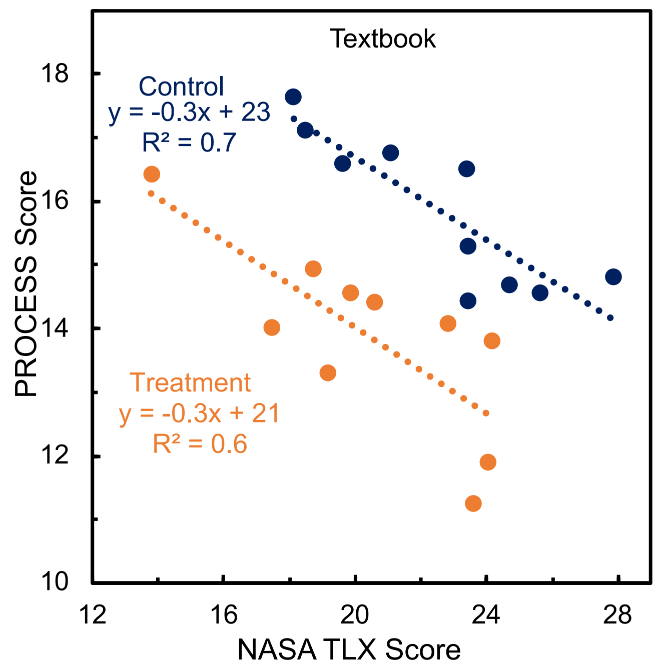
YouTube problems based on performance in each category of the problem-solving rubric. Individual scores in each category of the problem-solving rubric enabled for better understanding the areas of problem solving where participants faced the most challenges (Figure 3-18). Treatment group earned slightly higher PROCESS scores when solving YouTube problems than Textbook problems consistently in all components of the problem-solving rubric. For each component of the problem-solving rubric, students earned between 6 to 8% higher scores when students solve YouTube problems than when completing textbook problems (Table 3.9). Performance across all items of PROCESS for Textbook and YouTube problems possess no significant difference.

Table 3.9. Problem-solving exhibited when participants in the Treatment group complete 10 YouTube and 9 Textbook problems as a function of individual components of PROCESS rubric.

Problem solving components	Mean Textbook	Mean YouTube	p	Rank
Problem dentification	2.6 ± 0.3	2.8 ± 0.1	0.09	2
Representation	2.7 ± 0.3	2.9 ± 0.06	0.05	1
Organization	2.3 ± 0.3	2.4 ± 0.2	0.20	4
Calculation	2.1 ± 0.3	2.3 ± 0.1	0.21	5
Solution completion	2.5 ± 0.3	2.6 ± 0.2	0.18	3
Accuracy	1.8 ± 0.3	1.9 ± 0.2	0.37	6

First three components of the PROCESS (Problem identification, Representation, and Organization) represent conceptual skills. Whereas, the later three categories (Calculation, Solution completion, and Accuracy) measures analytical skills [59-61]. Considering conceptual skills displayed, students earned ~15% more in Problem identification and Representation than in the organization section (Figure 3-18 and Table 3.9).

On the other hand, PROCESS scores reveal that in the analytical section, students earned the lowest scores in Accuracy (~15% lower scores in Accuracy than in Calculation and 30% lower in Accuracy than in Solution completion). Overall, Organization, Calculation, and Accuracy were the only items were students did not display more than 80% of required skills in material and energy balance course. In fact, within the PROCESS as a whole, students displayed the least skill in Accuracy. Since Accuracy measures the final outcome of problem solving, lower scores in the Accuracy may result from missing or inaccurate steps associated with other components of the PROCESS. Here, Organization and Calculations component which were identified earlier as major lacking areas of the problem-solving and may be huge contributors the low Accuracy displayed by students.



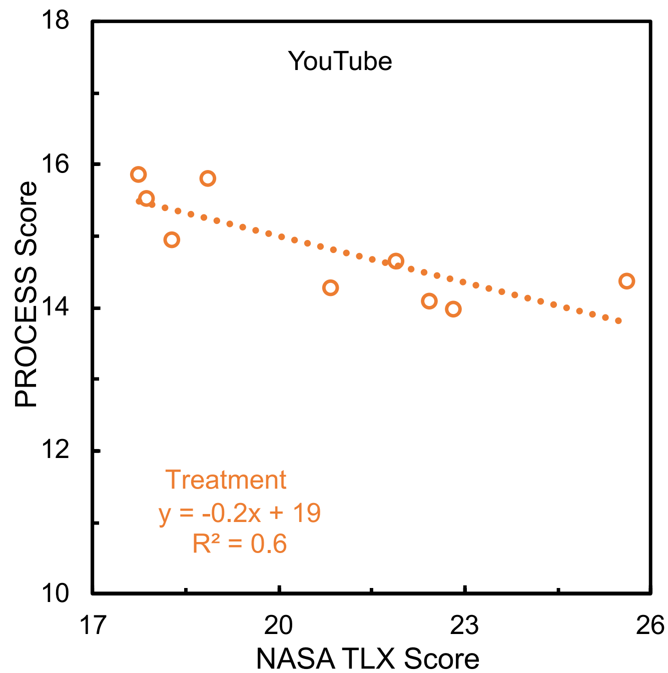


Figure 3-19: How students overall problem-solving ability correlate with problem difficulty when participants in control and treatment group solve 10 textbook problems (top), and when participants in treatment group solve 9 YouTube problems(bottom).

Generally, students tended to perform better when solving problems perceived to be less difficult and displayed lower abilities for problems they perceived to be more difficult denoted by negative signed slopes (Figure 3-19). Irrespective of group, similar trend and slope (-0.3) was observed when students solved Textbook problems portrayed in Figure 3-19 (top). However, when students in Treatment group solved YouTube problems, while a similar trend in the interaction between perception of difficulty and problem-solving ability, lower slope (-0.2) was measured (Figure 3-19 (bottom). When students complete Textbook problems, 10% (0.2 vs 0.3) more effect of perception of problem difficulty on problem-solving was measured than in YouTube problems. Perception of problem difficulty played a higher role in determining the success or the level of problem-solving ability displayed for Textbook problems unlike YouTube problems. Lesser dependency presented when solving YouTube problems may have resulted from the incorporation of videos into YouTube problems enabling students to visualize more and ease problem-solving process.

Table 3.10. CLASS scores measured in the beginning of course for Control group comprising responses from 23 students in year 1 and 27 students in year 2 study.

Categories	Year 1(%)	Year 2(%)	p
Overall	76 ± 14	75 ± 9	0.8
Real world connection	85 ± 21	85 ± 17	1.0
Personal interest	85 ± 18	90 ± 14	0.3
Sense making/effort	84 ± 19	80 ± 16	0.4
Conceptual connection	70 ± 25	60 ± 19	0.1
Applied conceptual thinking	59 ± 22	52 ± 19	0.2
Problem solving general	85 ± 16	81 ± 19	0.4
Problem solving confidence	82 ± 22	75 ± 23	0.3
Problem solving sophistication	73 ± 24	72 ± 23	0.9

Table 3.11. CLASS scores measured towards the end of course for Control group. CLASS scores consist of responses from 8 students in year 1 and 11 students in year 2 study.

Categories	Year 1(%)	Year 2(%)	p
Overall	89 ± 7	84 ± 8	0.2
Real world connection	97 ± 9	100 ± 0	0.3
Personal interest	98 ± 6	98 ± 5	1.0
Sense making/effort	95 ± 7	86 ± 21	0.3
Conceptual connection	92 ± 18	79 ± 25	0.2
Applied conceptual thinking	88 ± 21	79 ± 18	0.3
Problem solving general	94 ± 7	88 ± 13	0.3
Problem solving confidence	91 ± 13	82 ± 16	0.2
Problem solving sophistication	96 ± 8	88 ± 20	0.3

Table 3.12. Class scores measured in the beginning of course for treatment group comprising responses from 72 students in year 1 and 74 students in year 2 study.

Categories	Year 1(%)	Year 2(%)	p
Overall	77 ± 12	78 ± 10	0.6
Real world connection	88 ± 19	85 ± 19	0.3
Personal interest	90 ± 15	91 ± 16	0.7
Sense making/effort	85 ± 18	86 ± 15	0.7
Conceptual connection	72 ± 24	74 ± 23	0.6
Applied conceptual thinking	61 ± 24	60 ± 23	0.8
Problem solving general	89 ± 17	88 ± 13	0.7
Problem solving confidence	90 ± 19	84 ± 19	0.06
Problem solving sophistication	77 ± 27	75 ± 22	0.6

Table 3.13. CLASS scores measured towards the end of course for Treatment group comprising responses from 19 students in year 1 and 28 students in year 2 study.

Categories	Year 1(%)	Year 2(%)	p
Overall	81 ± 9	82 ± 10	0.7
Real world connection	93 ± 14	95 ± 12	0.6
Personal interest	95 ± 10	98 ± 5	0.2
Sense making/effort	88 ± 12	93 ± 9	0.1
Conceptual connection	82 ± 23	78 ± 24	0.6
Applied conceptual thinking	68 ± 25	65 ± 25	0.7
Problem solving general	91 ± 8	94 ± 8	0.2
Problem solving confidence	88 ± 15	91 ± 12	0.5
Problem solving sophistication	81 ± 19	82 ± 21	0.9

CLASS scores for participants in Treatment and Control group was measured in addition to PROCESS and TLX scores. Comparing CLASS responses for two years cohorts results in differences that are not statistically different evidenced by large p-values (Table 3.10, Table 3.11, Table 3.12, Table 3.13). Since differences for cohorts are not significant, scores were aggregated and used for analysis (Table 3.14, Table 3.15)

Table 3.14. CLASS scores measured in the beginning of course for 146 participants in the Treatment group and 50 participants in Control group over two years study.

Categories	Control (%)	Treatment (%)	p
Overall	75 ± 11	78 ± 11	0.10
Real world connection	85 ± 19	87 ± 19	0.5
Personal interest	87 ± 16	90 ± 16	0.3
Sense making/effort	82 ± 17	86 ± 17	0.2
Conceptual connection	65 ± 22	73 ± 24	0.04*
Applied conceptual thinking	55 ± 20	60 ± 24	0.2
Problem solving general	83 ± 18	88 ± 15	0.06
Problem solving confidence	78 ± 22	87 ± 19	0.006*
Problem solving sophistication	72 ± 24	76 ± 25	0.3

*Indicates a statistically significant difference.

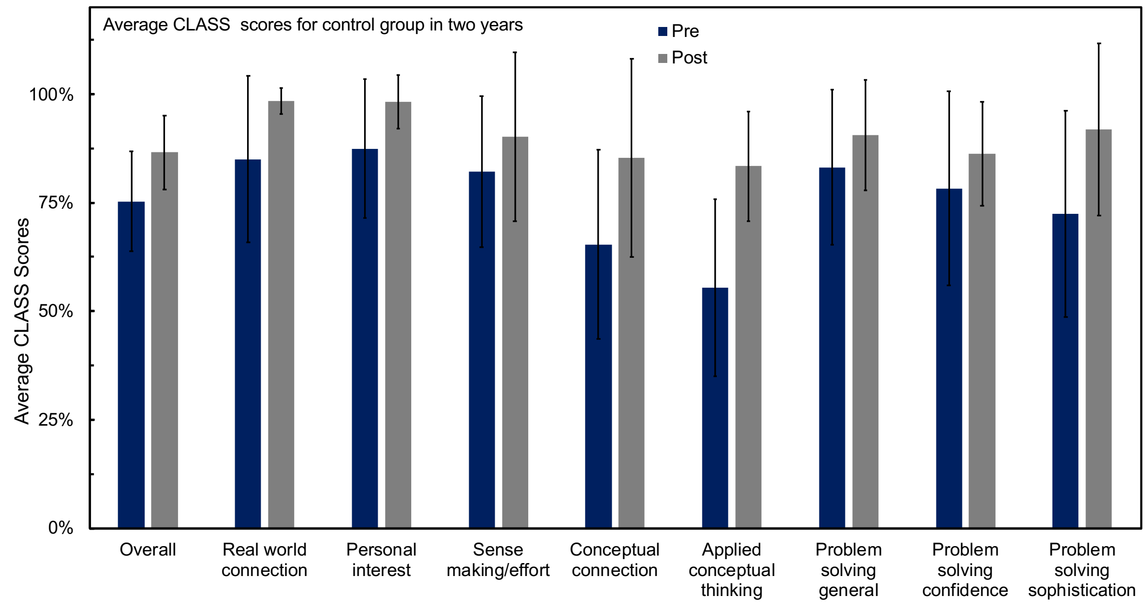
At inception of CLASS administration, ~200 scores were collected and in most categories of the CLASS, participants showed at least 70% agreement with desired attitudes and belief (Table 3.14). At the beginning, students in both groups displayed lower attitudes (mean CLASS score < 70%) towards conceptual connection and applied conceptual thinking than other categories. CLASS scores of Treatment group were slightly more aligned to expert beliefs than Control group. In most cases, attitudes of Control and Treatment group were statistically similar ($p > 0.05$). Significant attitude differences between groups were measured for conceptual connection and problem-solving confidence categories.

Table 3.15. CLASS scores measured towards the end of course for 47 participants in the Treatment group and 19 participants in Control group over two years study.

Categories	Control (%)	Treatment (%)	p
Overall	87 ± 9	81 ± 12	0.05
Real world connection	98 ± 3	94 ± 11	0.1
Personal interest	98 ± 6	96 ± 9	0.4
Sense making/effort	90 ± 20	91 ± 16	0.8
Conceptual connection	85 ± 23	80 ± 25	0.5
Applied conceptual thinking	83 ± 13	66 ± 17	0.0002*
Problem solving general	91 ± 13	92 ± 12	0.8
Problem solving confidence	86 ± 12	90 ± 15	0.3
Problem solving sophistication	92 ± 20	81 ± 21	0.06

*Indicates a statistically significant difference.

However, only about 66 participants responded to CLASS survey administered towards the end of course (Table 3.15). Towards the end of course, students mostly displayed attitudes that aligned to desired attitudes by at least 80%. Treatment group showed low agreement (CLASS score < 70%) with desired applied conceptual thinking attitudes. No significant difference was measured in all categories of CLASS besides in applied conceptual thinking ($p = 0.0002$)



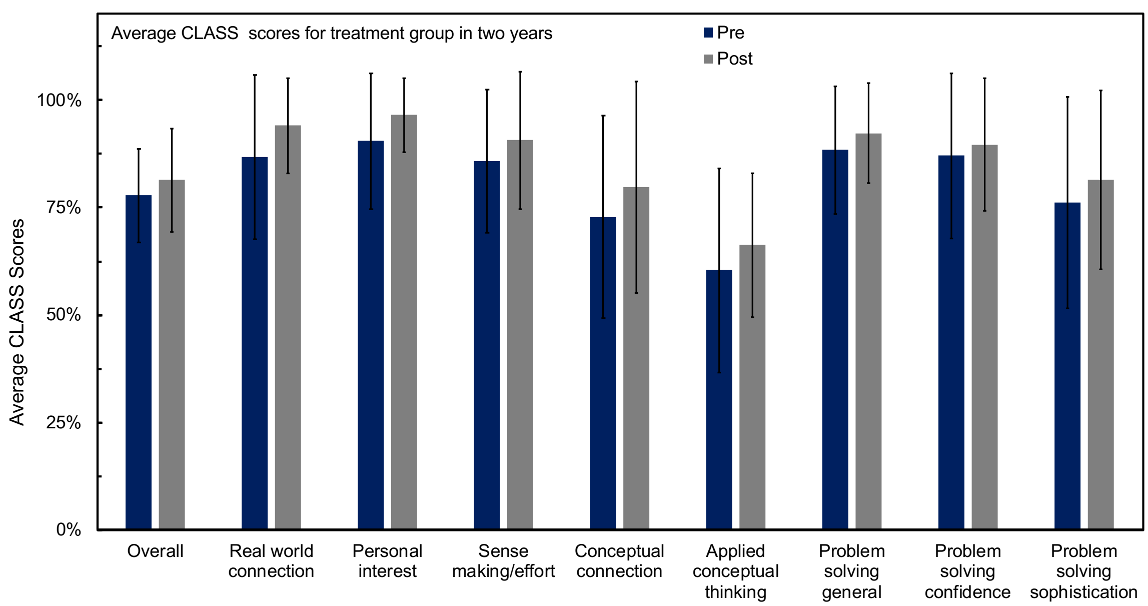


Figure 3-20: Average learning attitude of participants measured with CLASS for two years cohort study in the control (top) and treatment group (bottom).

Average CLASS scores collected for ~ 200 students over two years study show positive attitudes of students in material and energy balance course (Figure 3-20). At the beginning of the course, treatment group displayed 3% better overall attitudes towards learning in Chemical Engineering than control group. Treatment group exhibited ~10%

better attitudes in conceptual connection, applied conceptual thinking, and problem-solving confidence than students in the control group. Treatment group demonstrated between 2% to 6% more expert-like belief in real world connection, personal interest, sense making/effort, problem solving sophistication, and problem solving general than participants in the control group early in the study.

However, towards the end of the course, control group showed 6% higher overall learning attitudes and 26% more expert-like applied conceptual thinking than treatment group. Participants in the control group displayed between 2% to 13% more expert-like attitude in personal interest, real world connection, conceptual connection, and problem-solving sophistication than participants in the treatment group. While participants in the treatment group showcased a 1% to 4% more desired attitudes in sense making/effort, problem solving general and problem-solving confidence than the control group.

Table 3.16. Pre and post average CLASS scores for 50 participants in control group collected over two years study

Categories	Pre (%)	Post (%)	Normalized gain	Shift (%)	Effect size (g)
Overall	75 ± 11	87 ± 9	0.45	15	1.10
Real world connection	85 ± 19	98± 3	0.90	16	0.80
Personal interest	87 ± 16	98 ± 6	0.86	12	0.80
Sense making/effort	82 ± 17	90 ± 20	0.45	10	0.40
Conceptual connection	65 ± 22	85 ± 23	0.57	30	0.90
Applied conceptual thinking	55 ± 20	83 ± 13	0.63	50	1.50
Problem solving general	83 ± 18	91 ± 13	0.44	9	0.50
Problem solving confidence	78 ± 22	86 ± 12	0.37	10	0.40
Problem solving sophistication	72 ± 24	92 ± 20	0.70	27	0.90

Control group had a 15% positive shift in overall learning attitude measured within two semesters. Positive shifts in attitudes of about 30% or more was measured for conceptual connection, applied conceptual thinking and problem-solving sophistication for participants in the control group. Normalized gain calculations result in a positive gain by a factor 0.45 and a large effect size (Hedge's g = 1.1) in the overall learning attitudes of participants in the control group. Effect size calculation for most categories of the CLASS yielded Hedge's g > 0.7 indicating a large positive effect of material and energy balance instruction in the control group (Table 3.16).

Table 3.17. Pre and post CLASS scores for 200 participants treatment group collected over two years study.

Categories	Pre (%)	Post (%)	Normalized gain	Shift (%)	Effect size (g)
Overall	78 ± 11	81 ± 12	0.16	5	0.3
Real world connection	87 ± 19	94 ± 11	0.55	9	0.4
Personal interest	90 ± 16	96 ± 9	0.63	7	0.4
Sense making/effort	86 ± 17	91 ± 16	0.35	6	0.3
Conceptual connection	73 ± 24	80 ± 25	0.26	10	0.3
Applied conceptual thinking	60 ± 24	66 ± 17	0.15	10	0.3
Problem solving general	88 ± 15	92 ± 12	0.34	4	0.3
Problem solving confidence	87 ± 19	90 ± 15	0.20	3	0.2
Problem solving sophistication	76 ± 25	81 ± 21	0.22	7	0.2

Treatment group measured 5% positive shift in overall learning attitude assessed within one semester period of course administration. Positive attitude shift ~10%, was observed in real world connection, conceptual connection, and applied conceptual thinking. Overall learning attitude of participants in the treatment group was improved by a normalized gain factor of 0.16 with a small effect size (Hedge's g = 0.3). Effect size calculation in all categories of the CLASS resulted in a Hedge's g between 0.2 to 0.4. Positive effect size measured in the treatment group indicates that YouTube pedagogy did not show any negative influence for students learning Chemical Engineering (Table 3.17).

Chapter 4

Conclusion and Recommendation

In summary, homework-style, YouTube-inspired problems have been implemented in an undergraduate course in material and energy balances. YouTube problems were utilized as alternative homework problems for students and could be solved with or without videos. YouTube problems covered a wide variety of topics in material and energy balance course and can be adopted for a wide range of subject areas.

Here, a set of 9 YouTube problems in combination with 10 Textbook problems served as the basis for examining problem solving and rigor in this study. Through implementation of pseudo-control/treatment design, research examined impacts of replacing Textbook problems with YouTube problems. Research questions were directed towards evaluating rigor of YouTube problems, problem solving, and students' attitudes towards learning. Research utilized 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 rigor between YouTube and Textbook problems in responses for both first and second cohorts. Item analysis identified mental demand, effort, and frustration as the most significant factors to problem difficulty in solving material and energy balance problems. In contrast, 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 and problems did not require much physical exertion to complete.

An established rubric called PROCESS was revised to match the study and implemented to assess problem-solving ability across problems and groups. Analysis from PROCESS indicated better problem-solving skills for participants in the control group at inception of the study, suggesting group inequivalence in problem-solving ability. However, after YouTube intervention, the treatment and control groups exhibited nearly equivalent problem-solving skills on two problems. 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. Solving YouTube problems might be responsible for small effect size gains measured in the Treatment group within a 15-week semester unlike in the Control group where no gains were measured for over 30 weeks of instruction.

Item analysis within PROCESS identified Organization as the most difficult item within conceptual skill domain (Problem identification, Representation and Organization). Whereas, in the analytical skill section (Calculation, Solution completion, Accuracy), students scored lowest in the Accuracy stage. 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 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 better and aid problem-solving process.

In addition to assessing problem solving and difficulty, a validated survey known as CLASS measured attitudes towards learning. As with NASA TLX, consistent responses across two cohorts were found with CLASS. At the beginning and end of course, students reported positive attitudes towards learning engineering. At the beginning of course, lowest attitudes were reported for conceptual connection and applied conceptual thinking, which was improved towards the end of course. Overall, implementing YouTube problems as alternatives to Textbook problems exhibited no detrimental effect on the learning attitudes of students.

Students solution assessed with PROCESS rubric were totally selected through a random process. Many of the students selected did not turn in solutions to some of the homework problems. For example, in textbook problem TB1 (volume percent), only 8 out of 30 students in the Treatment group selected turned in the solution. To ensure that we make use of sufficient data, before students are selected, we ensure that students completed at least 80% of all the problems required for study. It may also be necessary after selection to check in order to ascertain what group (based on grades) of students were selected. Ensuring that selected students meet required minimum criteria would help us ensure that we are measuring outcomes of students who truly participated in the research intervention.

Current study utilized 9 YouTube problems, and by topic, represents 3 topics (reaction only, reaction with recycle, multiphase systems. 3 YouTube problems represents each of the 3 topics mentioned earlier. However, 10 Textbook problems implemented represented a wider range (8 topics/concepts). Rather than comparing an aggregate of scores for YouTube and Textbook problems, it may be necessary to compare Textbook and YouTube problems of very similar content and topic. For example, we may have 10 Textbook problems covering 5 topics. Two Textbook problems may represent each topic. Likewise, equal number of YouTube problems with similar contents with each of the 10 Textbook problems may be administered while NASA TLX surveys is collected and solutions are assessed with PROCESS. Matching problems of similar content will enable us make head to head comparison of problems and ensure that results from analysis is not influenced by difficulty of topics.

Since YouTube problems could be solved in the absence of the videos, future work may measure students' views on the videos. Embedding analytics features on video links will enable us to learn how students solve YouTube problems.

Future work may utilize control charts to identify students who indicate the highest difficulty with the NASA TLX survey. For example, NASA TLX responses over a two years period collected for each problem can establish control limits for each problem. With a control chart, attention will be drawn to participants who report mental workloads outside the control limits [63]. Control charts will help identify students' needs and may also lead further interviews where a fish-bone diagram could identify possible causes factors that influence students' perceptions [79].

Comparing completion rate of CLASS survey, we recorded over 50% drop of viable responses between pre and post surveys. However, giving some point incentives to students that complete surveys may help increase the number of responses. Besides, the issue of survey completion, another challenge with using CLASS is the possibility of measuring what student's think an ideal belief or attitude would be rather than what students believe [80]. Creating a section as part of the survey that requires students to provide an answer towards what they believe and what they think experts believe may enable us ensure that data collected are representative of students' beliefs and attitudes.

Aggregated analysis of CLASS for two years cohorts utilized scores from 200 students at the beginning and 70 students at the end of course. Besides non-completion of survey by participants, many CLASS responses were thrown away due to students failing to select the right choice in the attention check statement. In addition, future work may utilize a timer for the online CLASS surveys to ensure that participants are completing surveys faithfully. Using a timer may help collect cleaner data by discarding responses if the students complete survey within very short time. When running pre and post comparisons for CLASS, it might be important to utilize a matched data that only make use of scores for individuals who completed CLASS at the beginning and end of the course. A matched data will reduce noise by ensuring that shifts measured are due to changes in how students are thinking and not as a result of difference in the students who took the survey.

References

1. Internet World Stats. *Internet Usage Statistics; The Internet Big Picture: World Internet Users and 2018 Population Stats*. 2018 [cited 2019 January]; Available from: https://www.internetworldstats.com/stats.htm.

2. Kennedy, G.E., et al., *First year students' experiences with technology: Are they really digital natives?* Australasian journal of educational technology, 2008. **24**(1).

3. Roos, D., *How Net Generation Students Work.* 2007.

4. Lang, J.M., *Small Teaching: Everyday Lessons from the Science of Learning*. 2016: John Wiley & Sons.

5. Liberatore, M.W., *Improved student achievement using personalized online homework for a couse in material and energy balances.* Chemical Engineering Education, 2011. **45**(3): p. 184-190.

6. World Economic Forum, *Figure 10: Change in demand for core work-related skills, 2015-2020, all industries; Chapter 1: The Future of Jobs and Skills;*, in *The Future of Jobs: employment, Skills and Work force Strategy for the Fourth Industrial Revolution*. 2016. p. 21- 24.

7. Crimson. *Top 10 Jobs in 2030: Skills You Need Now to Land the Jobs of the Future: Future Skills*. 2018 [cited 2019 January]; Available from: https://www.crimsoneducation.org/us/blog/jobs-of-the-future.

8. Vest, C.M., *Infusing Real World Experiences into Engineering Education.* 2012.

9. Daigger, G.T., et al., *Real World Engineering Education Committee.* 2012.

10. Hawkins, J. and S. Blakeslee, *On intelligence*. 2005: St. Martin's Griffin.

11. Tasker, R. and R. Dalton, *Research into practice: visualisation of the molecular world using animations.* Chemistry Education Research and Practice, 2006. **7**(2): p. 141-159.

12. Wen, F. 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, 2016. **50**(3): p. 186-192.

13. Velazquez-Marcano, A., et al., *The Use of Video Demonstrations and Particulate Animation in General Chemistry.* Journal of Science Education and Technology, 2004. **13**(3): p. 315-324.

14. Liberatore, M.W., et al., *Student-created homework problems based on YouTube videos.* Chemical Engineering Education, 2013. **47**(2): p. 122-132.

15. Medina, J., *Brain Rules 12 Principles for Surviving and Thriving at Work, Home, and School*. 2008: Pear Press.

16. Brown, J.R. and M.B. McGrath, *Visual learning for science and engineering.* IEEE Comput Graph Appl, 2005. **25**(5): p. 56-63.

17. Al-Balushi, S.M. 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., 2014. **15**(1): p. 47-58.

18. Doidge, N., *The brain that changes itself: Stories of personal triumph from the frontiers of brain science*. 2007: Penguin.

19. Doidge, N., *The brain's way of healing: Remarkable discoveries and recoveries from the frontiers of neuroplasticity*. 2016: Penguin Books.

20. Freeman, S., et al., *Active learning increases student performance in science, engineering, and mathematics.* Proceedings of the National Academy of Sciences, 2014. **111**(23): p. 8410-8415.

21. Falconer, J.L., *Why not try active learning?* Aiche Journal, 2016. **62**(12): p. 4174-4181.

22. Arnaud, C.H., *Active Learning Beats Lectures.* Chemical & Engineering News, 2014. **92**(22): p. 31-31.

23. Cavanagh, S.R., *The Spark of Learning: Energizing the College Classroom with the Science of Emotion*. 2016: West Virginia University Press.

24. Liberatore, M.W., C.R. Vestal, and A.M. Herring, *YouTube Fridays: Student led development of engineering estimate problems.* Advances in Engineering Education, 2012. **3**(1): p. 1-16.

25. Liberatore, M.W., 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, 2017. **6**(1): p. n1.

26. Liberatore, M.W., *Active Learning and Just-in-time Teaching In a Material and Energy Balances Course.* Chemical Engineering Education, 2013. **47**(3): p. 154–160.

27. Liberatore, M.W., *High textbook reading rates when using an interactive textbook for a Material and Energy Balances course.* Chemical Engineering Education, 2017. **51**(3): p. 109-118.

28. Liberatore, M.W., *YouTube Fridays: Engaging the Net Generation in 5 Minutes a Week.* Chemical Engineering Education, 2010. **44**(3): p. 215-221.

29. Hadley, K.R. and K.A. Debelak, *Wiki technology as a design tool for a capstone design course.* Chemical Engineering Education, 2009. **43**(3): p. 194-200.

30. Edgar, T.F., *Information Technology and ChE Education: Evolution or Revolution?* Chemical Engineering Education, 2000(Fall): p. 290-295.

31. Abulencia, J.P., D.L. Silverstein, and M.A. Vigeant. *Cut! Adventures in Student-produced Instructional Videos for Thermodynamics*. in *ASEE Annual Meeting*. 2016. New Orleans, LA.

32. Liberatore, M., et al. *Student-Generated Problems that Reverse Engineer YouTube Videos*. in *ASEE Annual Conference proceedings*. 2018.

33. Bloom, B.S., ed. *Taxonomy of Education Objectives: the classification of educational goals*. 1956, Longsmans, Green: New York.

34. Anderson, L.W., *A critique of grading: Policies, practices, and technical matters.* education policy analysis archives, 2018. **26**: p. 49.

35. Watt, H.M., *Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia.* Educational studies in mathematics, 2005. **58**(1): p. 21-44.

36. Henderson, C., et al., *Grading student problem solutions: The challenge of sending a consistent message.* American Journal of Physics, 2004. **72**(2): p. 164-169.

37. Docktor, J.L., et al., *Assessing student written problem solutions: A problem-solving rubric with application to introductory physics.* Physical review physics education research, 2016. **12**(1): p. 010130.

38. D'Zurilla, T.J., A.M. Nezu, and A. Maydeu-Olivares, *Social problem solving: Theory and assessment.* 2004.

39. D'Zurilla, T.J. and A.M. Nezu, *Development and preliminary evaluation of the Social Problem-Solving Inventory.* Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1990. **2**(2): p. 156.

40. D'Zurilla, T.J. and A. Maydeu-Olivares, *Conceptual and methodological issues in social problem-solving assessment.* Behavior Therapy, 1995. **26**(3): p. 409-432.

41. Adachi, P.J. and T. Willoughby, *More than just fun and games: The longitudinal relationships between strategic video games, self-reported problem solving skills, and academic grades.* Journal of youth and adolescence, 2013. **42**(7): p. 1041-1052.

42. Safari, Y. and H. Meskini, *The effect of metacognitive instruction on problem solving skills in Iranian students of health sciences.* Global journal of health science, 2016. **8**(1): p. 150.

43. Özyurt, Ö., *Examining the Critical Thinking Dispositions and the Problem Solving Skills of Computer Engineering Students.* Eurasia Journal of Mathematics, Science & Technology Education, 2015. **11**(2).

44. Kourmousi, N., et al., *Validity and reliability of the Problem Solving Inventory (PSI) in a nationwide sample of Greek educators.* Social Sciences, 2016. **5**(2): p. 25.

45. Karaçam, A. and A. Pulur, *Identification the Relation between Active Basketball Classification Referees' Empathetic Tendencies and Their Problem Solving Abilities.* Universal Journal of Educational Research, 2016. **4**(8): p. 1912-1917.

46. Bélanger, C., et al., *The effect of cognitive-behavioral group marital therapy on marital happiness and problem solving self-appraisal.* The American Journal of Family Therapy, 2015. **43**(2): p. 103-118.

47. Bayindir Çevik, A. and N. Olgun, *Do Problem‐Solving Skills Affect Success in Nursing Process Applications? An Application Among T urkish Nursing Students.* International journal of nursing knowledge, 2015. **26**(2): p. 90-95.

48. Heppner, P.P., T.E. Witty, and W.A. Dixon, *Problem-Solving Appraisal and Human Adjustment: A Review of 20 Years of Research Using the Problem Solving Inventory.* The Counseling Psychologist, 2004. **32**(3): p. 344-428.

49. Heppner, P.P. and C.E. Baker, *Assessment in action: Application of the problem solving inventory.* Measurement and evaluation in counceling and development, 1997. **29**: p. 229-241.

50. Heppner, P.P. and C.E. Baker, *Applications of the problem solving inventory.* Measurement and evaluation in counseling and development, 1997. **29**(4): p. 229-241.

51. Shute, V.J., *Stealth assessment in computer-based games to support learning.* Computer games and instruction, 2011. **55**(2): p. 503-524.

52. Shute, V.J., et al., *Measuring problem solving skills via stealth assessment in an engaging video game.* Computers in Human Behavior, 2016. **63**: p. 106-117.

53. Shute, V.J. and M. Ventura, *Stealth assessment: Measuring and supporting learning in video games*. 2013: MIT Press.

54. Shute, V.J. and Y.J. Kim, *Formative and stealth assessment*, in *Handbook of research on educational communications and technology*. 2014, Springer. p. 311-321.

55. Montgomery, K., *Authentic tasks and rubrics: Going beyond traditional assessments in college teaching.* College teaching, 2002. **50**(1): p. 34-40.

56. Rosli, R., D. Goldsby, and M.M. Capraro, *Assessing students' mathematical problem-solving and problem-posing skills.* Asian social science, 2013. **9**(16): p. 54.

57. Benson, L., et al. *CU thinking problem solving strategies revealed*. in *ASEE Annual Meeting*. 2011. Vancouver, BC.

58. Grigg, S.J. and L. Benson. *Effects of Student Strategies on Successful Problem Solving*. in *ASEE Annual Meeting*. 2012. San Antonio, TX.

59. Grigg, S.J. and L. Benson. *Promoting problem solving proficiency in first year engineering process assessment*. in *ASEE Annual Meeting*. 2015. Seattle, WA.

60. Grigg, S.J. and L.C. Benson, *A coding scheme for analysing problem-solving processes of first-year engineering students.* European Journal of Engineering Education, 2014. **39**(6): p. 617-635.

61. Grigg, S.J., et al. *Process analysis as a feedback tool for development of engineering problem solving skills*. in *ASEE Annual Meeting*. 2013. Atlanta.

62. Lester, F.K., *Musings about mathematical problem-solving research: 1970-1994.* Journal for research in mathematics education, 1994. **25**: p. 660-660.

63. Grigg, S.J., S.K. Garrett, and L. Benson. *Using the NASA-TLX to Assess First Year Engineering Problem Difficulty*. in *IIE Annual Conference*. 2012.

64. Luger, S.K. and J. Bowles. *Two methods for measuring question difficulty and discrimination in incomplete crowdsourced data*. in *First AAAI Conference on Human Computation and Crowdsourcing*. 2013.

65. Thukral, D., et al., *DiffQue: Estimating Relative Difficulty of Questions in Community Question Answering Services.* arXiv preprint arXiv:1906.00145, 2019.

66. Vidulich, M.A. and P.S. Tsang. *Assessing subjective workload assessment: a comparison of SWAT and the NASA-bipolar methods*. in *Proceedings of the Human Factors Society Annual Meeting*. 1985. SAGE Publications Sage CA: Los Angeles, CA.

67. Reid, G.B. and T.E. Nygren, *The subjective workload assessment technique: A scaling procedure for measuring mental workload*, in *Advances in psychology*. 1988, Elsevier. p. 185-218.

68. Kilmer, K.J., R. Bateman, and D. Malzahn. *Techniques of subjective assessment: A comparison of the SWAT and modified Cooper-Harper scales*. in *Proceedings of the Human Factors Society Annual Meeting*. 1988. SAGE Publications Sage CA: Los Angeles, CA.

69. Bruning, R.H., G.J. Schraw, and R.R. Ronning, *Cognitive psychology and instruction*. 1999, Upper Saddle River, N.J.: Merrill.

70. Cao, A., et al., *NASA TLX: Software for assessing subjective mental workload.* Behavior research methods, 2009. **41**(1): p. 113-117.

71. Boser, U., *Learn Better: Mastering the Skills for Success in Life, Business, and School, or, How to Become an Expert in Just About Anything*. 2017: Rodale Books.

72. Grier, R.A., *How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores.* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2015. **59**(1): p. 1727-1731.

73. Hart, S.G. *NASA-Task Load Index (NASA-TLX); 20 Years Later*. in *Human Factors and Ergonomics Society Annual Meeting*. 2006.

74. Hart, S.G. and L.E. Staveland, *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, in *Human Mental Workload*. 1988. p. 139-183.

75. Sharek, D., *A Useable, Online NASA-TLX Tool.* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2011. **55**(1): p. 1375-1379.

76. Hendy, K.C., K.M. Hamilton, and L.N. Landry, *Measuring subjective workload: when is one scale better than many?* Human Factors, 1993. **35**(4): p. 579-601.

77. Pretz, J.E., A.J. Naples, and R.J. Sternberg, *Recognizing, Defining, and Representing Problems*, in *The Psychology of Problem Solving*, J.E. Davidson and R.J. Sternberg, Editors. 2003, Cambridge University Press: Cambridge, UK.

78. Adams, W.K., et al., *Modifying and validating the Colorado Learning Attitudes about Science Survey for use in chemistry.* Journal of Chemical Education, 2008. **85**(10): p. 1435.

79. Prastawa, H., N. Susanto, and T. Nofri. *Mental workload of undergraduate student (a case study in Industrial Engineering Department of Diponegoro University)*. in *SHS Web of Conferences*. 2018. EDP Sciences.

80. Gray, K.E., et al., *Students know what physicists believe, but they don't agree: A study using the CLASS survey.* Physical Review Special Topics-Physics Education Research, 2008. **4**(2): p. 020106.