The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
Paper ID #41043

ThermoVR: Using Virtual Reality and Playful Simulation to Teach and Assess Introductory Thermodynamics Concepts

David J Gagnon, Field Day Lab @ UW-Madison

David J. Gagnon is the research director of Field Day Laboratory at the University of Wisconsin - Madison. Field Day produces and researches educational video games that have won numerous awards (Serious Play, Meaningful Play, Public Media Awards,ASEE, and others) and are used by over a million students yearly in grades 4-20, across a diversity of subjects, from engineering to underwater archeology. Field Day is the organization behind Open Game Data, a national research infrastructure to explore the use of game data for educational research and learning engineering approaches for educational media.

Prof. John M. Pfotenhauer, University of Wisconsin, Madison

Professor John M. Pfotenhauer earned his BA, MA, and PhD degrees in physics from St. Olaf College and the University of Oregon in 1979, 1981, and 1984. For eight years he conducted research as part of the Applied Superconductivity Center at the Universit

Arganthael Berson, University of Wisconsin, Madison
Luke Swanson, University of Wisconsin, Madison

WIP: Developing a virtual laboratory for instruction and fine grained assessment of introductory thermodynamics

Importance

As an undergraduate course, introductory undergraduate thermodynamics is a foundational and ubiquitous course in engineering and science [1], including mechanical, chemical, and materials engineering, physics, and chemistry. In many departments, it functions as a "weed out" course, measuring a student's interest and proficiency in the field at an early stage in their program [2]. Jacobs & Freud [3] affectionately refer to introductory thermodynamics as an engineering student's "rite of passage."

Unfortunately, thermodynamics is a complex topic that is difficult for many students to understand and visualize [4], all too often leading to frustration and failure. These difficulties are evident to thermodynamics instructors: the topic is rich in concepts full of domain-specific vocabulary, and it usually requires the application of more than one fundamental principle or equation to analyze any given problem, confusing learners on how these equations interact [5]. Often, applying these principles requires navigating through large reference tables and diagrams because more familiar mathematical tools of calculus and vector analysis fail to deal with these phenomena [1]. Furthermore, Thermodynamics suffers from the same problem that plagues many engineering disciplines: students must not only understand the core concepts and processes within the discipline, but also master the mathematical tools required to solve numerical problems. For a large lecture course, this presents a significant issue for assessment. How is an educator able to formatively identify student misconceptions precisely, accurately and quickly when the topic is so complex?

The result is that a troublesome percentage of students fail the course or drop their engineering major because of the class. For example, a study across 55 offerings of introductory thermodynamics over eight years and nine instructors at the University of Texas, San Antonio, reports that only 52.7% of course enrolments led to completion, with subsequent attempts having lower rates than initial attempts [6].

In this Work in Progress paper, we describe a project to develop a fine-grained assessment of core thermodynamic concepts embedded within an interactive simulation environment known as a virtual laboratory.

Virtual Laboratories for Thermodynamics Education

Virtual laboratories play an increasingly important role in undergraduate engineering education, especially following the COVID-19 shutdowns of many in-person laboratory experiences. Virtual laboratories are digital learning environments that allow learners to conduct investigations using simulated material and apparatus. They have been studied extensively in science and engineering education [7]. They are increasingly utilized in educational settings, online learning, and training in industry as appropriate alternatives to physical laboratories [8]. Virtual laboratories offer advantages to traditional laboratories in terms of logistics, including lower costs and less setup time [9], as well as the advantages all online software have for learning, such as scalability, increased access, and near-zero distribution costs.

Regarding instructional efficacy, many well-controlled comparison studies report no differences between physical and virtual laboratories [7]. For example, Wiesner and Lan [10] compared virtual and physical equipment for measuring heat exchange, mass transfer, and humidification. They found no differences in the resulting performance of chemical engineering students in terms of underlying engineering principles. Ma and Nickerson [11] reviewed 39 studies comparing hands-on, simulated, and remote laboratories in engineering education, finding no differences in their educational effectiveness.

Additional studies describe the unique affordances of virtual laboratories for engineering education. For example, several studies illustrate the advantages of virtual, interactive explorations of unobservable phenomena compared with physical experiments of observable phenomena. For example, university students who investigated simulated electric circuits showing moving electrons acquired more conceptual knowledge than those using hands-on materials [12]. Similarly, students using virtual optics materials displaying light rays outperformed those using physical materials [13]. Studies show virtual experiments can enable students to use complex inquiry practices to separate variables that might be difficult to use in physical experiments [14], [15]. There is also the idea that virtual experiments are well suited to developing conceptual knowledge because, unlike physical instruments, they produce "clean" data. For example, first-year chemistry students using virtual experiments performed better than those using a typical laboratory on conceptual understanding measures, partly attributable to the messy data produced by the physical lab [16]. Finally, Zacharia et al. [9] found that virtual laboratories offered students more time to experience an experiment and concentrate on its conceptual aspects than the corresponding physical laboratories because the virtual laboratories allowed faster manipulation of the materials involved in the experiments of the study's curriculum.

Current Practices for Assessing Thermodynamics Understanding

As all educators know, assessing student understanding is critical to offering appropriate educational interventions. Unfortunately, the topic of thermodynamics, like many engineering disciplines, contains a complex interdependence between qualitative, conceptual understandings of the various phenomena and the quantitative, mathematical processes required to solve problems. In an attempt to disentangle, previous educators have developed several assessment instruments focused on the conceptual understandings and problem-solving approaches required by thermodynamics. One of the original assessment instruments is known as the Thermodynamics Concept Inventory (TCI; See Figure 1) [17] which created a pre/post measure for introductory thermodynamics to compare students before and after taking courses, instructors' performance, and different institutions. The two-part Thermodynamic Concept Survey (TCS) [18] is a similar work. The first section focuses on temperature and heat transfer, while the second concentrates on the first law of thermodynamics and process. Other known instruments include the Thermodynamic Diagnostic Test (TDT) [19] and the Heat and Energy Concept Inventory (HECI)[20].

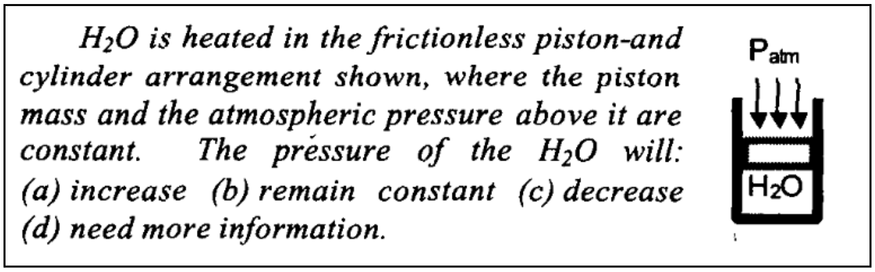
H_2O is heated in the frictionless piston-and cylinder arrangement shown, where the piston mass and the atmospheric pressure above it are constant. The pressure of the H_2O will: (a) increase (b) remain constant (c) decrease (d) need more information.

Figure 1. Example Item from the Thermodynamics Concept Inventory.

Each of these assessments follow the multiple-choice structure. While this approach is certainly compatible with relatively low effort grading, it is not an optimal design. The first problem is that most evidence for student conceptual thinking is outside of the assessment input, the option selected, often seen as jottings on paper. It is therefore the full responsibility of the grader to develop ways to interpret, score, and provide feedback on student work beyond their multiple-choice selection, if that information is even available. Second, grading and reporting back to the students takes time, and additional work is required to aggregate individual student scores into any insights about common misconceptions. It may be days or weeks before the results of the assessments are communicated to the students, and instructors may never be supported to see the larger patterns so they can adjust their instructional strategies.

Approaches for Improving Assessment

Embedded assessment has gained prominence as educators and researchers recognize the limitations of conventional assessment methods in accurately measuring student learning and providing timely feedback. The approach is rooted in the principles of formative assessment and authentic assessment, emphasizing the importance of assessing students in the context of their learning activities [21], [22].

Conceptually, embedded assessment is an educational approach that integrates learner evaluation within the learning process, aiming to measure and support student learning in real-time. This method stands in contrast to traditional forms of assessment which are typically administered separately from instruction. Embedded assessment involves the integration of assessment tasks into instructional activities, making the assessment process less intrusive and more reflective of students' actual learning processes [23]. Assessment tasks are designed to be directly relevant to the learning objectives and often require students to apply their knowledge and skills in authentic contexts. This approach enables educators to assess not only the final product of learning but also the learning process itself, including students' problem-solving strategies, critical thinking, and ability to apply knowledge in real-world situations [24].

Embedded assessment comes with many challenges. Teachers must be skilled in designing assessment tasks and in interpreting the evidence of learning these tasks provide [25]. Due to embedded assessment's inherently qualitative approach, it does not naturally scale to many students easily. Finally, unlike standardized tests, which are designed to ensure consistency and comparability, embedded assessments are highly contextualized and may vary from one learning environment to another. Establishing robust criteria and rubrics can help address these concerns, but it requires ongoing effort and collaboration among educators [26].

Evidence-Centered Design (ECD) provides a systematic approach to principled assessment design that is compatible with an embedded context [27] but also is compatible with virtual learning environments and automated analysis [28]. ECD is grounded in the idea that assessments should be designed around a model of student cognition and learning in specific content areas, and it emphasizes the collection of evidence to support claims about students' knowledge, skills, and competencies. ECD also supports the creation of assessments that are fair and accessible to all students by considering the diversity of learners and the contexts in which they demonstrate their knowledge and skills.

The ECD framework consists of three interrelated models that provide assessment validity by design [29]. The **student model** defines the specific knowledge, skills, and abilities that the assessment aims to measure. It outlines the content area and the types of tasks or situations in which these abilities will be demonstrated. The **evidence model** specifies what evidence is needed to support claims about a student's knowledge, skills, or abilities. It involves identifying observable variables or response patterns that indicate the presence of the underlying competencies being assessed. Bayesian network models are commonly employed for this analysis. The **task model** describes the tasks or situations that will elicit the evidence defined in the evidence model. It includes specifications for designing tasks that are aligned with the domain model and that will generate the necessary evidence to support claims about student abilities. Together, these models define an explicit relationship between what a student does within a specific context, and the claims that can be made about their thinking.

ECD is compatible with digital learning environments, offering the speed of computationally grading simple multiple choice exams with the richness provided by embedded assessment. Instead of developing questions for students to answer, assessment designers ask the question: "What situations can I create that would elicit learner thinking to be demonstrated in an observable way?' For example, in the context of a digital simulation, ECD provides designers with a way to capture meaningful attributes about how a learner interacts with a simulation problem, defining how each move (e.g. changing a parameter of the simulations input) provides evidence for (or against) student understanding of some component of the system. This is interesting for our problem of assessing thermodynamics because it provides a method to operate at fine grain sizes (assessing individual concepts separate from the mathematics), while still scaling to large class sizes easily because student interaction data is easily processed by a computer (processing the relevance of each student action using a explicit evidence model).

Virtual Laboratories for Thermodynamics Assessment

While there are examples of using games, simulations and virtual laboratories for teaching thermodynamics concepts [30], [31], [32], little work is well known that utilize these approaches for assessment. Expanding to other disciplines, et al. [33] developed a virtual laboratory for teaching and assessing the construction of electronic circuits. The laboratory afforded students a breadboard and various connectable components that had the potential to simulate a large number of projects. Following an ECD-inspired approach, the assessment utilized the actions performed by the students, what they call virtual behavior observation, as evidence for assessing their abilities, knowledge, and understanding. They used Bayesian networks to model these constructs and included other components, such as prior knowledge and interface familiarity, to explore alternate explanations for evidence other than student proficiency.

Project Goals

ThermoVR is a thermodynamics virtual lab for undergraduate instructional purposes. The laboratory simulates an experiment with water contained in a piston cylinder device. The user can affect the thermodynamic state of the water by acting on the piston-cylinder device in various ways (e.g. heat, cool or insulate the system, increase or decrease pressure, fix the piston in place, etc.) and the simulation responds in terms of a real time visualization of the piston-cylinder device, numerical outputs as well as a 3d plot of the current thermodynamic state. This project builds on prior work to develop ThermoVR [32] and now focuses on developing features that support fine-grained, scalable assessment.

Goal 1: Develop a mechanism to prompt students to interact with the ThermoVR simulation in particular ways that provide opportunities to assess students' conceptual understanding of the critical relationships, attributes, and processes for change of five fundamental properties of thermodynamics (pressure, temperature, volume, entropy, internal energy) across the key regions (liquid, vapor dome, superheated vapor) for water. Note this goal focuses on qualitative concepts, not the mathematical machinery required to solve numerical problems.

Goal 2: Develop mechanisms for student-facing reports that support their own self-assessment and allow them to focus their study on specific areas.

Goal 3: Develop mechanisms for instructors to understand their students both individually and on aggregate, so they can intervene as they see fit.

The Conceptual Assessment Framework for ThermoVR

To develop the next version of ThermoVR, which will guide student activity and create opportunities for students to demonstrate their understanding of thermodynamics principles, we leverage Evidence Centered Design and develop what Mislevy [27] refers to as a Conceptual Assessment Framework (CAF).

The Task Model - Defining Observable Student Behaviors in Constructed Contexts. For ThermoVR, each task is a collection of prompts that integrate instructional prompts with assessment items, seen within the system as a "lab." The player is able to select a lab to guide their interactions using a tablet computer within the virtual environment (Figure 2). The tablet then loads the lab and prompts students with various steps they will follow to perform the laboratory experience, such as configuring the apparatus to prepare for the experiment, calling attention to elements of the simulation they should notice, providing questions for reflection, providing "sandbox" opportunities to freely experiment, and defining open-ended challenges to use the tools of the system to achieve particular outcomes. This tablet device is fully integrated with the simulation and is the readout for virtual instruments. It can monitor and advance when a pre-programmed target state is reached (e.g., reaching 300 degrees Kelvin) within the simulation, what we call a "check." These prompts and checks coalesce into the steps that make up the lab activity. In terms of the overall ThermoVR project, these labs are being authored and reviewed by a community of thermodynamics instructors across several midwestern engineering colleges, and are aligned to the topics in their collective syllabi and thermodynamics textbooks.

Some prompts within a given lab activity are purely instructional, do not produce evidence for student thinking, and simply instruct the student to perform an action and take notice of the result. The prompts are grouped together into short sequences of training and are designed to teach students about the interface and thermodynamics concepts. These prompts follow the format:

- Use the [specific tool] to …
- Record the current [pressure/temperature/volume, entropy, enthalpy] value
- Observe [some dynamic element of the simulation]

A second category of prompts, however, operates as an embedded assessment and creates evidence for student thinking. These tasks require students to judge based on their understanding of the underlying system and use the tools provided to achieve a goal. Assessment items come in the following formats:

- Increase/Decrease [pressure/temperature/volume] using any tool
- Move to a [pressure/temperature/volume] of [value] using any tool
- Create a constant [pressure/temperature/volume, entropy] condition
- Move to the [liquid/vapor dome/vapor/superheated vapor] region

When an assessment item is the current step in the lab, the learner receives notification that the game will evaluate their next move as a formative assessment. Therefore, while failure is possible (with very low consequences), they should focus on achieving their assigned "task." The assessment notification will help to improve validity.

The student's final work product is a path describing all the points reached within the thermodynamic space during the activity (i.e., particular values of pressure, temperature, volume, entropy, and internal energy) and the actions performed by the student to achieve that path. Some parts of this path are obtained by following the prompts. Other path components require students to use the supplied tools at their discretion to reach certain conditions. The user receives a visual representation of their path: a line drawn on the 3d thermodynamic graph (see Figure 2) and a completed checklist on the tablet. This product becomes a collection of time-series telemetry data events stored in a database for analysis.

Figure 2: Image of a prototype for ThermoVR version 2, with the tablet device on the far left, a piston-cylinder apparatus in the top middle, controls in the bottom center, and the 3d thermodynamic graph on the lower right

The Student Model - Defining Variables to Describe Components of Student Thinking. This assessment will focus on building a model of student understanding of three primary constructs. These constructs are particularly useful for describing student conceptual knowledge and informing instruction. They were identified by a review of engineering education literature and the experiences of the ThermoVR co-PI and co-design community.

Given the nature of the thermodynamics simulator, student actions are effectively limited to nine controls. Each control represents one physical tool that is plausible in the piston-cylinder apparatus, such as insulation, a heating element and a weight to place on the piston. Each of the tools available within the simulation will have some effect on the current state. This effect is mediated by which (if any) of the thermodynamic properties are being held constant (eg. constant pressure) and the current region of the state (eg. superheated vapor). Therefore, different tools have distinct effects based on the current region of the state and any constant conditions that are at play. For example, if the system is in a state of constant pressure and in the liquid region, using the heater tool will increase the temperature, until the system reaches the 2-phase/vapor dome region at a barrier called the saturated liquid line. Upon crossing the saturated liquid line, the state is now in the vapor dome region and temperature will remain constant even as heat is added, unless the system is first put into a constant volume scenario using the clamp tool.

Together, understanding the role of each tool in terms of thermodynamic principles, methods for creating constant conditions and the properties of each of the regions and boundaries are critical to achievement in the practices of thermodynamics engineering as they allow the student to break down complex problems into a series of individual steps. Our complete student model (See Figure 3) attempts to describe these components of understanding by defining a total of 26 individual student model variables, which inform three primary constructs of thermodynamics properties and the first law of thermodynamics.

> Construct 1: Student understanding of the influence of the tools and physical components on the thermodynamic state. 5 student variables describe student understanding of how each of the physical tools is related to the thermodynamic state of the system.
>
> Construct 2: Student understanding of physical conditions that create constant thermodynamic pathways. 4 student variables describe student ability to create steady pressure, volume, temperature and entropy conditions.
>
> Construct 3: Student understanding of the unique properties of each region. 17 student variables will describe student understanding of how the thermodynamic properties of water are influenced by each of the 4 regions, under each of the 4 constant conditions. An additional variable will describe student understanding of the saturation line that contains the vapor dome.

For consistency, we construct each variable name from concatenation of the construct (e.g. “TOOL”), followed by the secondary (i.e “-pressure”) and tertiary attributes (i.e “-vapor_dome”) if present. This results in variables such as “TOOL-clamps”, “CONSTANT-pressure-liquid” and “PROPERTY-superheated_vapor-constant_volume”

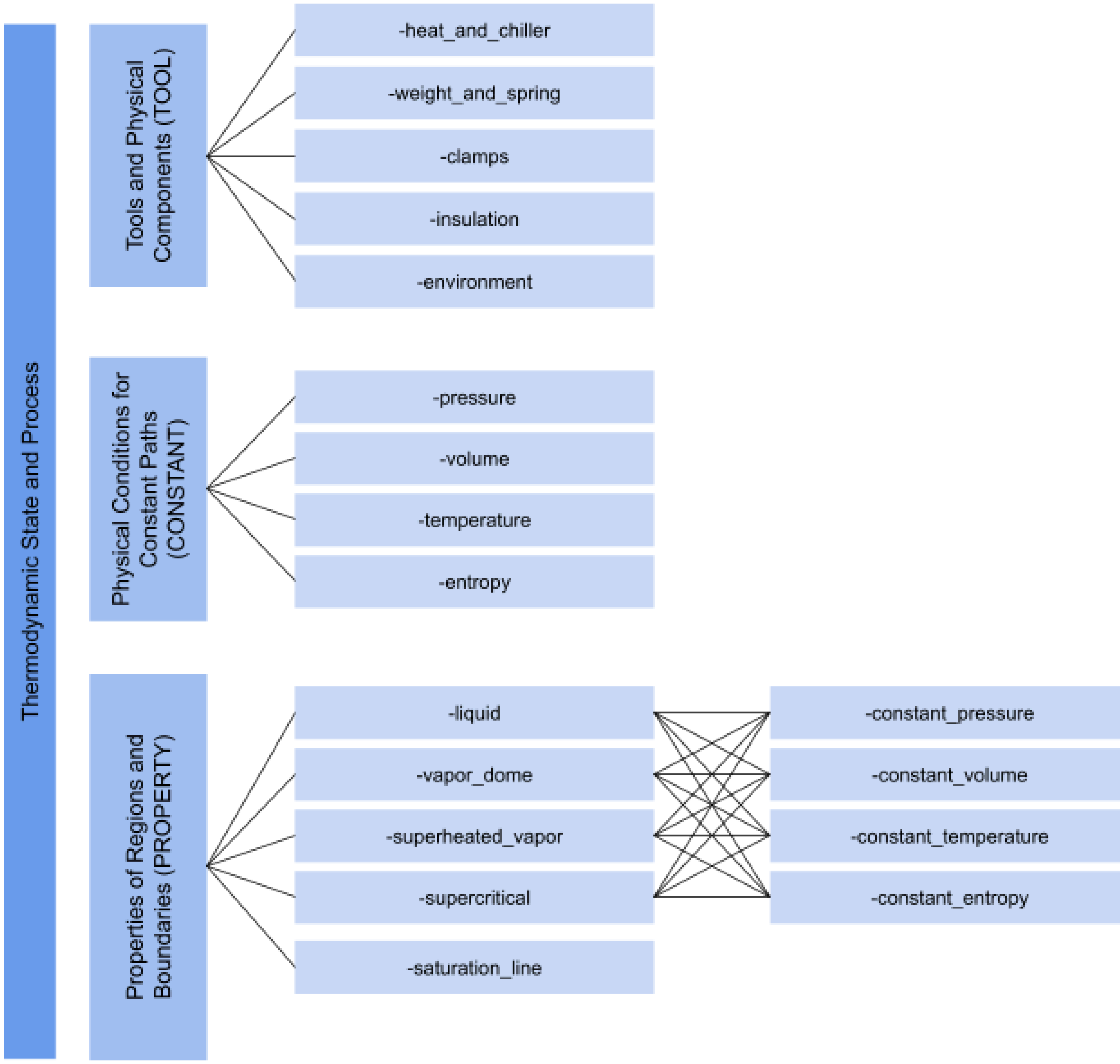

Figure 3: Visualization of Components of Thermodynamic State to be Assessed

Evidence model - How student actions are interpreted to make claims. The evidence model describes how different student actions are interpreted when performing a task. In the ThermoVR assessment, the evidence model is a simple algorithm that is followed whenever an assessment task is provided to the student, informed by the context of the task type (create a constant condition / change a thermodynamic property), the current region, and any constant conditions, resulting in updates to one or more student model variables.

Continuing from the example above, if the water is currently in a liquid state under constant pressure and the task asks the player to "Increase the temperature of the water to 100 degrees C," if the players next move is to use the heater tool, we have evidence that they understand the effect of the heater concerning temperature, as well as how the temperature changes in the liquid under constant pressure. As such, we add to the *TOOL-heater_and_chiller* and *PROPERTY-liquid-pressure.* If the task prompt did not inform the player that a constant pressure condition was active, and therefore the player needed to determine that fact themselves, we also add to *CONSTANT-pressure-liquid.* In this way, we use a collection of scoring tables to represent how the evidence model should interpret each player action following a task prompt, and if any player model variables should be updated. For the above tasks, the scoring is somewhat complex as each of the tasks requires multiple skills, therefore more than one student model variable is updated. On the other hand, simple tasks such as asking the student to "add heat energy to the system" only require one student skill and therefore only one student model variable is updated.

Unfortunately, a direct line of inference is not available between most observable actions and claims regarding their unobservable skills. For the example above, where the learner made a productive move, one interpretation is that they successfully utilized both prerequisite skills, but an alternate interpretation is that they randomly picked the correct tool, a not-so-unlikely one in nine chance. Similarly, if they were to make an unproductive move, such as adding a weight on top of the piston, we cannot be sure which part of the requisite skills they did not possess, or if it was simply an accidental move. Therefore, each of the student model variables is a latent variable which must be probabilistically estimated.

To develop these statistical claims, we use a machine learning approach known as Bayesian Knowledge Tracing (BKT) [34]. BKT has been used and theorized extensively by the educational assessment community [35] and in particular by artificial intelligence researchers as student models for intelligent tutoring systems [36]. BKT models individual student skills as a probability that they will demonstrate mastery on the next opportunity, based on their previous performances and the following four parameters:

- **Probability of Prior Knowledge ($p(L_0)$)**: This parameter represents the probability that the student knew the skill before any interaction with the system.
- **Probability of Transit ($p(T)$)**: This parameter measures the probability the student learns the skill after attempting a problem related to that skill.
- **Probability of Guess ($p(G)$)**: This parameter accounts for the likelihood that the student guesses the answer correctly without actually knowing the skill. It helps distinguish between true knowledge and lucky guesses.
- **Probability of Slip ($p(S)$)**: The slip parameter is the probability that the student, despite knowing the skill, incorrectly answers a problem. This could be due to mistakes, misunderstandings, or other factors unrelated to their actual knowledge level.

Each of these parameters must be initially estimated for each student model variable. For ThermoVR, we will use an initial dataset of historical learner performances to provide starting values for each prior knowledge $p(L_0)$ and transit $p(T)$, then update these values at a later date as more learner data is available. Guess $p(G)$ and slip $p(S)$ values are more easily estimated based on the design of the ThermoVR interface itself and the problem being attempted. For example, given that the interface only provides nine parameters to either raise or lower from the current value, the player always has a 1 in 18 chance of guessing the correct move, even if they did not understand the concept.

In the first step following each assessment task, the BKT approach will calculate a conditional probability of learning $\mathbf{p(L_t)}$, for the current task:

If the response is **correct**, the conditional probability of learning is:

$$p\left(L_{t+1} | obs = correct\right) = \frac{p\left(L_t\right) * (1 - p(S))}{p\left(L_t\right) * (1 - p(S)) + (1 - p\left(L_t\right)) * p(G)}.$$

If the response is **incorrect**, the conditional probability of learning is:

$$p\left(L_{t+1} | obs = incorrect\right) = \frac{p\left(L_t\right) * p(S))}{p\left(L_t\right) * p(S) + (1 - p\left(L_t\right)) * (1 - p(G)}.$$

Based on this calculation, BKT then updates the probability that the student has now learned the skill p(L_{t+1}):

$$p\left(L_{t+1}\right) = p\left(L_{t+1} | obs\right) + \left(1 - p\left(L_{t+1} | obs\right)\right) \times p(T).$$

To understand how these estimations are changing over time, we record the individual updates to each p(L) student model variable. This data can be visualized as a moment-by-moment learning curve [38], a plot of the changes to student model variables in each BKT model on the vertical axis over the count of performances considered on the horizontal axis. This analysis not only provides us with a simple slope calculation to show performance trends over time, but indications of exactly when during the educational experience the player demonstrated evidence of learning.

Capturing, Processing and Reporting Analysis from ThermoVR

While a full description of a data telemetry system are well outside the scope of this paper, the high-level infrastructure and approaches are easily described. For this work we adopt and extend the Open Game Data research infrastructure [39] which provides opensource technologies and conventions for logging "telemetry" data from the ThermoVR system, storing those signals as time-series data, then processing those data into usable descriptions of a specific student's performance. The logging is initiated within the ThermoVR code, which sends one *event* to a logging server for each action taken by the user (e.g. selecting the upper clamp tool and pulling to a maximum volume of 5L), feedback given by the system (e.g. notifying the player that the state crossed a region boundary), or progression event that takes place within the activity (e.g. completing a lab activity). Each of these events contains metadata such as an anonymous user identification code, timecode, versioning of the application, and various state variables about the player and simulation. On VR versions of ThermoVR, high frequency position and rotation values for the head and each hand are also logged. Together these events constitute a time-series description of the learner interaction, and can be used for analysis or fed back into ThermoVR to create a full replay of the original interaction for qualitative methods.

Taking this stream of events as input, a collection of *features* are calculated. Features are calculated values that aggregate many events into one value that describes a single use of the ThermoVR tool, what we call a *session*, or an aggregate of all of a given user's sessions into a single *user* metric, or across many users into a *population* metric. A metric can be as simple as active_duration or level1_attempts, or more complicated such as an array of values that describe each change to each student model variable.

These calculated features are made available to a web-based reporting system via requests made to a RESTful API. The reporting system provides a simple interface to select specific sessions, users or populations and visualize their features.

Ongoing Research Plan

In Fall of 2023, the initial version of the ThermoVR system [32], which demonstrated the core simulation and interface features was piloted with small numbers of students in six North American universities' engineering courses. Surveys and interviews were conducted with the students and their instructors and the results were used to identify and correct technical issues and develop insights into the interest and needs of these audiences.

In Spring of 2024 a second usability pilot and small-scale evaluation was conducted with 9 students at the UW-Madison campus, using the new ThermoVR version containing the structured laboratory activities and embedded assessment data telemetry. An instructor-created, multiple-choice instrument was developed using items adapted from the Thermodynamics Concept inventory and previous quiz and examination questions. Students were given this as an online pretest, followed by a 30 minute intervention. A post test of the same items was administered followed by a semi-structured interview developed to elicit data about the usability of the virtual laboratory. The pilot demonstrated the validity of the approach and technical readiness for deployment in a full study.

Following one additional round of development based on the pilot, an evaluative study will be conducted in Fall of 2024 with approximately 180 participants measuring student learning gains and perceived usage of the assessment reporting. Following a similar protocol as above, the students will also be given the report from this assessment. The instructor will be given an aggregate report for all students who used the system. Both students and instructor will report on their perceptions of ThermoVR as a formative assessment tool and how the reports aided their teaching and learning goals.

Conclusion

In this paper we describe a work in progress to develop a new assessment for use in introductory thermodynamics courses, a critical opportunity for supporting the next generation of engineers. While much prior assessment work has been done into disentangling the concepts of thermodynamics topics from the mathematical tools required to solve numerical thermodynamics problems, the format of the resultant assessments, multiple choice tests, is still problematic. Drawing from recent advances in assessment design and utilizing machine learning approaches such as BKT, we have described an assessment tool that is embedded within a virtual laboratory activity. This assessment promises to describe student skills in a fine grained, 26 dimensional report using data collected by student interactions with a digital simulation of a piston-cylinder system. The assessment report can be provided in realtime to the learner to help them direct study efforts as well as aggregated over an entire class to inform an educator about how to allocate instruction time. While the project is technically mature, studies are still underway with evaluative pilot studies taking place in February, 2024 and a full evaluation planned for Fall of 2024.

Thinking forward, we hope this paper outlines a novel approach for teaching and assessing student thinking in the context of a simulated virtual laboratory environment that could be applied to engineering disciplines outside of thermodynamics. Beyond our own implementations and forthcoming evaluation efforts, we hope to see ongoing research to explore the efficacy of various design elements of virtual laboratories in context of various students' needs and strengths. While certainly not every student should study engineering, it is our responsibility to give those that make an attempt to have the best possible learning experience.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. (DUE #2142103). This work also leverages technical infrastructure known as Open Game Data which was supported by the Learning Agency TOOLS competition and the National Science Foundation under Grant No. (DRL #1907384).

References

[1] D. Probst and Y. Zhang, "A Gentle Bridge between Dynamics and Thermodynamics," in *2013 ASEE Annual Conference & Exposition Proceedings*, Atlanta, Georgia: ASEE Conferences, Jun. 2013, p. 23.48.1-23.48.6. doi: 10.18260/1-2--19062.

[2] R. D. Manteufel, "A Spiral Approach for Mechanical Engineering Thermodynamics," in *Proceedings of the ASME International Mechanical Engineering Conference*, Nashville, TN, 1999.

[3] T. Jacobs, J. Caton, J. Froyd, and K. Rajagopal, "Redesigning the First Course of Thermodynamics to Improve Student Conceptualization and Application of Entropy and Second Law Concepts," in *2014 ASEE Annual Conference & Exposition Proceedings*, Indianapolis, Indiana: ASEE Conferences, Jun. 2014, p. 24.1033.1-24.1033.17. doi: 10.18260/1-2--22966.

[4] N. Mulop, K. Mohd-Yusof, and Z. Tasir, “A Review on Enhancing the Teaching and Learning of Thermodynamics,” *Procedia - Soc. Behav. Sci.*, vol. 56, pp. 703–712, Oct. 2012, doi: 10.1016/j.sbspro.2012.09.706.

[5] C. Vieira, A. J. Magana, R. E. García, A. Jana, and M. Krafcik, “Integrating Computational Science Tools into a Thermodynamics Course,” *J. Sci. Educ. Technol.*, vol. 27, no. 4, pp. 322–333, Aug. 2018, doi: 10.1007/s10956-017-9726-9.

[6] A. Karimi and R. Manteufel, “Correlation of Prerequisite Course Grades with Student Performance,” in *2013 ASEE Annual Conference & Exposition Proceedings*, Atlanta, Georgia: ASEE Conferences, Jun. 2013, p. 23.343.1-23.343.15. doi: 10.18260/1-2--19357.

[7] T. de Jong, M. C. Linn, and Z. C. Zacharia, “Physical and Virtual Laboratories in Science and Engineering Education,” *Science*, vol. 340, no. 6130, pp. 305–308, Apr. 2013, doi: 10.1126/science.1230579.

[8] M. del R. Uribe, A. J. Magana, J.-H. Bahk, and A. Shakouri, “Computational simulations as virtual laboratories for online engineering education: A case study in the field of thermoelectricity: COMPUTATIONAL SIMULATIONS AS VIRTUAL LABORATORIES,” *Comput. Appl. Eng. Educ.*, vol. 24, no. 3, pp. 428–442, May 2016, doi: 10.1002/cae.21721.

[9] Z. C. Zacharia, G. Olympiou, and M. Papaevripidou, “Effects of experimenting with physical and virtual manipulatives on students’ conceptual understanding in heat and temperature,” *J. Res. Sci. Teach.*, vol. 45, no. 9, pp. 1021–1035, Nov. 2008, doi: 10.1002/tea.20260.

[10] T. F. Wiesner and W. Lan, “Comparison of Student Learning in Physical and Simulated Unit Operations Experiments,” *J. Eng. Educ.*, vol. 93, no. 3, pp. 195–204, Jul. 2004, doi: 10.1002/j.2168-9830.2004.tb00806.x.

[11] J. Ma and J. V. Nickerson, “Hands-on, simulated, and remote laboratories: A comparative literature review,” *ACM Comput. Surv.*, vol. 38, no. 3, p. 7, Sep. 2006, doi: 10.1145/1132960.1132961.

[12] N. D. Finkelstein *et al.*, “When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 1, no. 1, p. 010103, Oct. 2005, doi: 10.1103/PhysRevSTPER.1.010103.

[13] G. Olympiou, Z. Zacharias, and T. deJong, “Making the invisible visible: enhancing students’ conceptual understanding by introducing representations of abstract objects in a simulation,” *Instr. Sci.*, vol. 41, no. 3, pp. 575–596, May 2013, doi: 10.1007/s11251-012-9245-2.

[14] Klahr, David, Triona, Lara, Strand-Cary, Mari, and Siler, Stephanie, “Virtual vs. Physical Materials in Early Science Instruction: Transitioning to an Autonomous Tutor for Experimental Design,” in *Beyond Knowledge: The Legacy of Competence*, J. Zumbach, N. Schwartz, T. Seufert, and L. Kester, Eds., Dordrecht: Springer, 2008, pp. 163–172. [Online]. Available: https://doi.org/10.1007/978-1-4020-8827-8_23

[15] K. W. McElhaney and M. C. Linn, “Investigations of a complex, realistic task: Intentional, unsystematic, and exhaustive experimenters,” *J. Res. Sci. Teach.*, vol. 48, no. 7, pp. 745–770, Sep. 2011, doi: 10.1002/tea.20423.

[16] K. Pyatt and R. Sims, “Virtual and Physical Experimentation in Inquiry-Based Science Labs: Attitudes, Performance and Access,” *J. Sci. Educ. Technol.*, vol. 21, no. 1, pp. 133–147, Feb. 2012, doi: 10.1007/s10956-011-9291-6.

[17] K. C. Midkiff, T. A. Litzinger, and D. L. Evans, “Development of Engineering Thermodynamics Concept Inventory instruments,” in *31st Annual Frontiers in Education Conference. Impact on Engineering and Science Education. Conference Proceedings (Cat. No.01CH37193)*, Reno, NV, USA: IEEE, 2001, pp. F2A – F23. doi: 10.1109/FIE.2001.963691.
[18] P. Wattanakasiwich, P. Taleab, M. D. Sharma, and I. D. Johnston, “Development and Implementation of a Conceptual Survey in Thermodynamics,” *Int. J. Innov. Sci. Math. Educ.*, vol. 21, no. 1, pp. 29–53, 2013.
[19] C. Kamcharean and P. Wattanakasiwich, “Development and Implication of a Two-tier Thermodynamic Diagnostic Test to Survey Students’ Understanding in Thermal Physics,” *Int. J. Innov. Sci. Math. Educ.*, vol. 24, no. 2, pp. 14–36, 2016.
[20] M. Prince, M. Vigeant, and K. Nottis, “Development of the Heat and Energy Concept Inventory: Preliminary Results on the Prevalence and Persistence of Engineering Students’ Misconceptions,” *J. Eng. Educ.*, vol. 101, no. 3, pp. 412–438, Jul. 2012, doi: 10.1002/j.2168-9830.2012.tb00056.x.
[21] P. Black and D. Wiliam, “Assessment and Classroom Learning,” *Assess. Educ. Princ. Policy Pract.*, vol. 5, no. 1, pp. 7–74, Mar. 1998, doi: 10.1080/0969595980050102.
[22] G. Wiggins, “The Case for Authentic Assessment,” *Pract. Eval.*, vol. 2, no. 2, 2019, doi: https://doi.org/10.7275/ffb1-mm19.
[23] L. A. Shepard, “The Role of Assessment in a Learning Culture,” *Educ. Res.*, vol. 29, no. 7, pp. 4–14, 2000.
[24] L. Darling-Hammond and F. Adamson, “Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning”.
[25] M. Heritage, “Formative Assessment: What Do Teachers Need to Know and Do?,” *Phi Delta Kappan*, vol. 89, no. 2, pp. 140–145, Oct. 2007, doi: 10.1177/003172170708900210.
[26] W. J. Popham, “Assessment Literacy for Teachers: Faddish or Fundamental?,” *Theory Pract.*, vol. 48, no. 1, pp. 4–11, Jan. 2009, doi: 10.1080/00405840802577536.
[27] R. J. Mislevy, R. G. Almond, and J. F. Lukas, “A BRIEF INTRODUCTION TO EVIDENCE-CENTERED DESIGN,” *ETS Res. Rep. Ser.*, vol. 2003, no. 1, pp. i–29, Jun. 2003, doi: 10.1002/j.2333-8504.2003.tb01908.x.
[28] R. J. Mislevy *et al.*, “Psychometrics and game-based assessment,” in *Technology and testing: Improving educational and psychological measurement*, Routledge New York, 2016, pp. 23–48.
[29] R. J. Mislevy, “Validity by Design,” *Educ. Res.*, vol. 36, no. 8, pp. 463–469, Nov. 2007, doi: 10.3102/0013189X07311660.
[30] J. M. Pfotenhauer, D. J. Gagnon, M. J. Litzkow, and C. C. Blakesley, “Designing and Using an On-line Game to Teach Engineering,” in *FIE2009*, San Antonio, TX, Jul. 2009, pp. 1–5.
[31] J. M. Pfotenhauer, D. J. Gagnon, M. J. Litzkow, and C. Pribbenow, “Game Design and Learning Objectives for Undergraduate Engineering Thermodynamics,” in *2015 ASEE Annual Conference & Exposition*, Seattle, WA, Jun. 2015, pp. 1–14. doi: 10.18260/p.24147.
[32] J. M. Pfotenhauer and D. J. Gagnon, “ThermoVR: A Virtual Laboratory to Enhance Learning in Undergraduate Thermodynamics,” presented at the ASEE Virtual Annual Conference, 2021. [Online]. Available: https://peer.asee.org/37908
[33] I. E. Achumba, D. Azzi, V. L. Dunn, and G. A. Chukwudebe, “Intelligent Performance Assessment of Students’ Laboratory Work in a Virtual Electronic Laboratory Environment,” *IEEE Trans. Learn. Technol.*, vol. 6, no. 2, pp. 103–116, Apr. 2013, doi:

10.1109/TLT.2013.1.
[34] A. T. Corbett and J. R. Anderson, “Knowledge tracing: Modeling the acquisition of procedural knowledge,” *User Model. User-Adapt. Interact.*, vol. 4, no. 4, pp. 253–278, 1995, doi: 10.1007/BF01099821.
[35] M. J. Culbertson, “Bayesian Networks in Educational Assessment: The State of the Field,” *Appl. Psychol. Meas.*, vol. 40, no. 1, pp. 3–21, Jan. 2016, doi: 10.1177/0146621615590401.
[36] M. C. Desmarais and R. S. J. D. Baker, “A review of recent advances in learner and skill modeling in intelligent learning environments,” *User Model. User-Adapt. Interact.*, vol. 22, no. 1–2, pp. 9–38, Apr. 2012, doi: 10.1007/s11257-011-9106-8.
[37] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon, “Individualized Bayesian Knowledge Tracing Models,” in *Artificial Intelligence in Education*, vol. 7926, H. C. Lane, K. Yacef, J. Mostow, and P. Pavlik, Eds., in Lecture Notes in Computer Science, vol. 7926. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 171–180. doi: 10.1007/978-3-642-39112-5_18.
[38] R. S. Baker, A. Hershkovitz, L. M. Rossi, A. B. Goldstein, and S. M. Gowda, “Predicting Robust Learning With the Visual Form of the Moment-by-Moment Learning Curve,” *J. Learn. Sci.*, vol. 22, no. 4, pp. 639–666, Oct. 2013, doi: 10.1080/10508406.2013.836653.
[39] D. J. Gagnon and L. Swanson, “Open Game Data: A Technical Infrastructure for Open Science with Educational Games,” in *Serious Games*, vol. 14309, M. Haahr, A. Rojas-Salazar, and S. Göbel, Eds., in Lecture Notes in Computer Science, vol. 14309. , Cham: Springer Nature Switzerland, 2023, pp. 3–19. doi: 10.1007/978-3-031-44751-8_1.