5-4-2021

Considering different research-based activities by using a k-means cluster analysis

Amin Bayat Barooni
Georgia State University

Considering different research-based activities by using a k-means cluster analysis

by

Amin Bayat Barooni

Under the Direction of Brian D. Thoms, PhD

ABSTRACT

A variety of activities are commonly used in college physics courses including lab, tutorials, and studio curricula. Instructors must choose among using research-based activities, designing their own activities or modifying existing activities. Instructors' choices depend on their own goals and the goals of activities from which they are choosing. To assist them in developing or modifying activities for their situation, we examine research-based activities to determine their goals and the features of the activities associated with these goals. Since most activities ask students to perform tasks to assist them in learning, sixty-six activities from eleven different research-based curricula were coded for student actions. The coding scheme containing 49 codes in ten categories was developed from a subset of activities, interviews with some of the activity designers, and recommendations from the American Association of Physics Teachers 2014 lab report. The results were examined using k-means cluster analysis revealing three design clusters. We label these clusters Thinking like a Scientist, Learning Concepts, and Building Models. These three clusters reflect diverse design goals. In the Thinking like a Scientist cluster, activities emphasize design of experiments by students, discussion, error analysis, reasonableness checking, supporting claims, and making assumptions or simplifications. The Learning Concepts cluster focuses on prediction of results and experimental observations. The Building Models cluster emphasizes discussion and answering physics or math questions that do not use collected data. This work connects common features appearing in physics activities with the goals and strategies of the designers. In this way it may provide instructors with a more straightforward way to create activities which achieve their desired outcomes.

INDEX WORDS: Coding, interview, AAPT lab report 2014, k means cluster analysis, research-based physics activities, research-inspired physics activities.

Considering different research-based activities by using a k-means cluster analysis

by

Amin Bayat Barooni

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2021

Considering different research-based activities by using a k-means cluster analysis

by

Amin Bayat Barooni

Committee Chair:	Brian D. Thoms
Committee:	Megan Connors
	D. Michael Crenshaw
	Sumith Doluweera
	Sidong Lei

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2021

DEDICATION

I want to dedicate this thesis to my family, who always support and motivate me to learn—also, my advisors Dr. Brian D. Thoms, Dr. Jashua Von Korff.

ACKNOWLEDGMENTS

The author gratefully acknowledges financial support from the National Science Foundation Award No. 1347510. Special thanks to my advisors Dr. Brian D. Thoms and Dr. Joshua Von Kor, for all these years of support and guidance. I thank the designers who gave their valuable time to participate in this study. I also thank Jon Gaffney, Jacquelyn Chini for their helpful suggestions, and Monica Cook, Zeynep Topdemir, Hannah Pamplin, Myat Pho, Ibraheem Robins, Brian Ferguson, and Kyle Simmons for their collaboration during the research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
Overview and background

1.1 Brief review of the dissertation

In this dissertation, Chapter 1 will review what Physics Education Research (PER) is and how physics education researchers investigate data by using qualitative, quantitative, and mixed research methods. Moreover, I will explain the procedure of coding data in qualitative research by using an example from my research. Finding agreements among researchers is a critical process that I will explain by utilizing inter-rater reliability (IRR) and an example from my research. Chapter 2 is designated for reviewing previous research projects as constructions for my final research and introduction of dissertation research with research questions. Chapter 3 will explain how the data was collected using interviews with designers of design groups, research-based activities, and recommendations for the Undergraduate Physics Laboratory Curriculum 2014 by the AAPT Committee on Laboratories. I will expand each section and review how this procedure results in reliable data with cycles of collecting data, coding, and IRR. Furthermore, I will explain how we analyze qualitative data to convert it to quantitative data. Chapter 4 explains the use of k-means cluster analysis to achieve a reliable result. In Chapter 4 I also explain the different clusters resulting from the k-means cluster analysis, discuss their characteristics, and use interviews and literature to support these results. Also, I will dig more into results to find design strategies for each design group. Each research has its limitations. I will talk about it in the last sections of the Chapter 4. Chapter 5 is designated for follow-up research. I will explain how research-based activities use different representations to assist students in achieving their design goals. Also, I will talk more about how specific activities follow strategies that cause them to concentrate more than others on designing and evaluating experiments. Chapter 6 discusses future directions and explaining applications for my research.

1.2 Physics Education Research (PER)

Physics Education Research (PER) concentrates on investigating students' learning physics at all levels and improving teaching procedures to support students with diverse backgrounds (1). According to McDermott (2), there is evidence that PER can enhance student knowledge in physics. The traditional point of view about teaching physics is that it is an art and not a science (3). PER researchers use empirical research methods that are used in applied science. They document their procedures and results and conduct regular searches on how students from different levels learn scientific skills (2). They apply the results to conduct instructional materials' development and evaluate their effectiveness based on students learning (2). This demonstrates that teaching can be considered a science, not an art (2). Using PER methods, student difficulties can recognize, investigate, and efficiently address by systematic research process, curriculum designing, and instruction (2). The main concentration of PER is to investigate "student as a learner, rather than on the instructor as a teacher." (2) PER researchers conduct different studies among the diverse population of students to achieve this goal (2).

There are two principal characteristics of investigation methodologies in PER, qualitative research and quantitative research. Each has diverse concentrations, and each responds to different issues (4).

1.3 Qualitative research in PER

According to Beichner, a physics education researcher selects this type of research when he/she decides to investigate small populations, or limited data (4). Because information extracted from qualitative data is deep with details, it is not efficient to select extensive samples, so researchers choose small samples and investigate data deeply (4). On the other hand, because researchers have small samples, they cannot generalize results (4). It is not easy to transcribe and analyze hours of interviews and recorded data (4). Qualitative research is an umbrella expression for various research approaches such as grounded theory, ethnographic, narrative research, case studies, and phenomenology (5). In qualitative research, the researcher collects data through contact with people or their documents (5). Participant observations and in-depth interviewing are two standard methods to collect data (5).

For example, for participant observation the researcher participates in observing the people and taking notes, such as recording student activities in a class. If there is an observation protocol, then he/she uses it to record data. The researcher collects data from these observations and finds common features among the data. The researcher should do this process frequently to make sure he/she can record as much data in the field notes as possible.

If the researcher's goal is to extract data by interview, he/she must follow specific steps.

The first step is to contact the interviewee, explain the research risk, and sign the consent form. Researcher then construct an interview protocol in which they plan the questions he/she should ask and how to lead the conversation. For example, a list of questions about interviewee's background or perspective about a specific topic. The best questions are open-ended questions. The interviewer should not ask unnecessary questions during the interview and give the interviewee this chance to expand on his/her ideas. The researcher should then run the interview process by asking questions, recording the interview, and writing critical points on paper to generate more questions and lead the conversation.

Document analysis is another type of qualitative research that was mainly used in my research. These documents can include transcripts of interviews, observation fields note, pictures, video or audio records, written documents such as physics activities. For analyzing documents, qualitative researchers should use the coding process.

1.3.1 Coding process

According to Charmaz (6), "Coding means categorizing segments of data with a short name that simultaneously summarizes and accounts for each piece of data." For every researcher, codes present how they choose, break, and rearrange data and start analytic computing (6). Coding is crucial since it connects data collection to theory. The first step of coding is initial coding (6). Codes show the researcher's point of view since "we choose the words that constitute our codes" and "we define what we see as significant in the data and describe what we think is happening (6)." If researchers persist too much on preexisting ideas, they force the data to fit them instead of making their codes fit the data (6). Two techniques of

II. TESTING EXPERIMENT: COVERING THE IMAGE

Your friend Joshua suggests the following hypothesis: "The fraction of the image that is visible in a plane mirror depends on how much of the mirror is covered." Design an experiment to test Joshua's hypothesis.

Available equipment: Plane mirror on stand, object, masking tape, paper, whiteboard, meter stick.

Design and describe the experiment that you plan to perform. Remember that your prediction of the outcome of the experiment must follow from the hypothesis you are testing. Then perform the experiment and record the outcome. Explain the outcome using a ray diagram. Discuss whether the outcome agrees or disagrees with the prediction. If it disagrees, how would you convince Joshua that his hypothesis has been disproven?

Figure 1.1 ISLE, Physics 194, Lab 8: Reflection and Mirrors (Taken from reference 10)

initial coding are "Word-by-Word," and "Line-by-Line" codings (6). For many researchers, Line-by-Line coding is the first step of coding (6). One of the remarkable hints about coding is we should code actions (6). The next step is to find common features among code and categorize them in the category (6). There are two critical practices during the coding process (6):

1) The constant comparative method is a critical analytic distinction.

2) Memo writing is a vital step.

In my research, I analyzed documents for student actions, not themes or structures. The following example explains the coding process used in my research for a section of Investigative Science Learning Environment (ISLE), Physics 194, Lab 8: Reflection and Mirrors as illustrated in Figure 1.1 (7) .

The first step is open coding which consists of freely coding segments of data, inventing new codes, and comparing segments of data with one another. We coded student actions mentioned in the activities. Table 1.1 shows how this process works. After we find codes in different documents and compare them to each other.

Text	Code
Design and describe the experiment that you plan to perform.	Students Design an experiment They describe their plan
Then perform the experiment and record the outcome.	Students record their outcome
Explain the outcome using a ray diagram.	Students draw diagram to explain their outcome
Discuss whether the outcome agrees or disagrees with the prediction.	Students compare their results with prediction
If it disagrees, how would you convince Noelle that her hypothesis has been disproven?	Students should explain their idea

Table 1.1 First step of the coding process

We repeat the process to make sure that there are no new common codes. Then we categorize them in categories in a process know as axial coding (relating the codes to one another to develop categories) (8). We use this code book for analyzing new documents. There are different methods for registering the codes, such as printing the documents and using markers, using software, and using spreadsheets. We selected the spreadsheet method since we want to analyze our codes quantitatively after the qualitative research. Table 1.2 shows how we register data in the spreadsheet. We use the star sign to show that we observe the code in the specific unit cell.

Category	Code	ISLE Lab 8_Section II
Written Representation: Different kind of written representations	Written Word	*
	Student Chosen Representation	*
	Multiple Choice	*
	Diagram	*
Prediction: The process of prediction an experiment or making hypothesis by students and comparing the result of the experiment with the prediction	Make Prediction	*
	Check Prediction	*
Observation: The process of collecting data by observation and using it	Observe Data from Equipment	*
	Use Observed Data from Equipment	*
Design: The process of design, improvement, making hypothesis for an experiment or math procedure by students	Procedure Design	*

Table 1.2 Registering codes in the spreadsheets for using in quantitative analysis

1.3.2 Inter-rater reliability (IRR)

Each researcher has a different point of view about qualitative data, therefore they each must find a procedure that decreases their biases as much as possible. They do this process to make sure that when another researcher repeats it, he/she approximately finds the same results. The process is known as inter-rater reliability (IRR). It is the process that two or more raters (researchers) agree. It concerns consistency in the implementation of a rating system. Researchers use different statistics to estimate IRR, for example, percentage agreement, kappa, product–moment correlation, and intraclass correlation coefficient. High IRR rates are assigned to a high degree of agreement among researchers. Low IRR rates refer to a low degree of agreement (9).

As illustrated in Table 1.3, cells a and d designate sequentially the numbers of codes that both researchers reported (or both did not report). Cells b (c) shows the numbers of codes researcher 1 (2) reported but researcher 2 (1) did not. The number of times that researcher 1 reported (did not report) the codes is f_1 (f_2). Similarly, the number of times that researcher 2 report (did not report) the code is g_1 (g_2) (10). The total of the number of times reported and not reported, that is, the maximum number of opportunities to report, is n.

We define P_0 as the proportional agreement observation,

$$P_0 = \frac{a+d}{n}$$

and P_c as the proportional agreement expectation,

$$P_c = \frac{f_1}{n} \times \frac{g_1}{n} + \frac{f_2}{n} \times \frac{g_2}{n}.$$

In our research, two different researchers code the same activity and compare their agreements with Cohen's kappa coefficient to measure correlation among themselves. According to Vach (11), Cohen's kappa is a "standard tool for the analysis of agreement on a binary outcome between two observers or two methods of measurement." The mathematical definition of Cohen's kappa is (10),

Category	Code	Rater 1	Rater 2
Written Representation: Different kind of written representations	Written Word	*	*
	Student Chosen Representation	*	
	Multiple Choice	*	
	Diagram	*	*
Prediction: The process of prediction an experiment or making hypothesis by students and comparing the result of the experiment with the prediction	Make Prediction	*	*
	Check Prediction	*	*
Observation: The process of collecting data by observation and using it	Observe Data from Equipment	*	*
	Use Observed Data from Equipment	*	
Design: The process of design, improvement, making hypothesis for an experiment or math procedure by students	Procedure Design	*	*

Table 1.3 An example of two raters observations

$$\kappa = \frac{P_0 - P_c}{1 - P_c}.$$

To explain how we use Cohen's kappa in our research, we'll use the data in Table 1.2 and expand it for two different raters (Rater 1 and Rater 2). Suppose these two raters code the ISLE, Physics 194, Lab 8: Reflection and Mirrors, section II as illustrated in Figure 1.1 and want to measure how much they agree about observing codes in that unit cell. First, both of them, without sharing their idea, code the activity then insert data in the separate spreadsheets that contain the codebook. After that, they have a meeting and merge their data like Table 1.3.

Then they add numbers of observing codes in Table 1.4. As you see, both raters report five common codes in the coding unit cell. Rater 1 finds two codes that Rater 2 could not find them. Rater 2 finds one code that Rater 1 could not find it. Finally, both of them could not find one code. Now if we use formulas for P_0 and P_c to calculate kappa.

$$P_0 = \frac{a+d}{n} = \frac{5+1}{9} = 0.67,$$

$$P_c = \frac{f_1}{n} \times \frac{g_1}{n} + \frac{f_2}{n} \times \frac{g_2}{n} = \frac{6}{9} \times \frac{7}{9} + \frac{3}{9} \times \frac{2}{9} = 0.67 \times 0.78 + 0.33 \times 0.22 = 0.52 + 0.07 = 0.59.$$

Then according to Kappa formula:

$$\kappa = \frac{P_0 - P_c}{1 - P_c} = \frac{0.67 - 0.59}{1 - 0.59} = \frac{0.08}{0.41} = 0.2.$$

I will explain more about how to interpret Kappa numbers in Chapter 3.

1.4 Quantitative research in PER

Quantitative research is selected by a physics education researcher when he/she decides to investigate extensive samples. In contrast to qualitative research, quantitative research does allow generalization from data. The disadvantage of this method is that the researcher

		Rater 2		Total
		Report	Not Report	
Rater 1	Report	a 5	b 2	$g_1 = a + b$ 7
	Not Report	c 1	d 1	$g_2 = c + d$ 2
Total		$f_1 = a + c$ 6	$f_2 = b + d$ 3	$n = g_{1+}g_2$ 9

Table 1.4 Comparing repetitions of codes reported by two different raters (Taken from Ref. 11)

can not interpret data for individual students (4). Researchers usually can't get enough data about the interactions among inviduals (such as students and instructors) to apply quantitative methods. In the quantitative investigation, researchers often use assessment tests or surveys in a multiple-choice arrangement that can be performed on a larger scale. There is a list of reliable assessment tests on the Physport website (12). Researchers can use these assessment tests to investigate vast numbers of students. Because of their considerable statistical power, researchers can generalize results.

1.5 Mixed research methods in PER

If a physics education researcher needs to analyze data deeply, he/she typically uses qualitative research for analyzing small samples. However, he/she then can't generalize the results. On the other hand, if a researcher needs to generalize results, then he/she should use quantitative research, but he/she needs a large sample and cannot analyze data deeply. To have statistical power to generalize results and access some in-depth information about data, researchers can use a mixed method approach combining both qualitative and quantitative research (4). It is the method that we used for our study. In the mixed research method, the researcher first uses qualitative research to collect data, then he/she find codes and categories in documents to convert them to a sufficient amount of quantitative data to apply mathematical and statistical methods.

CHAPTER 2

Reviewing previous and current research

2.1 Introduction

In our research, we aim to support instructors to use the features of different research-based activities to design their own activities. In the early stages of this work, we interviewed designers of research-based activities and analyzed their activities. We found a specific design strategy known as "revisiting cycles" and published our results in PERC 2016 proceeding paper (13). We followed our research by interviewing more designers and analyzing more research-based activities. We also analyzed the AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum 2014 (14). Again, we found "revisiting cycles" as a theme and explained how designers' perspectives about conceptual learning and "thinking like a physicist" form their design features. The result of this middle stage of the research was published in PERC 2017 proceeding paper (15). In the later stage of the research, we analyzed more research-based activities and used quantitative analysis (k-means cluster) to find general patterns of similarities and differences among research-based activities. This final stage revealed three design strategies used by physics curriculum developers. I will discuss details of that work in Chapters 3-5. In this chapter we discuss the background of each stage of the research; Revisiting strategy, Thinking like a physicist and Conceptual learning, and Three design strategies revealed by k-means cluster.

2.2 Revisiting strategy

In the first stage of the research that we published in Physics Education Research Conference (PERC 2016) as a proceeding paper (13), we examined three curricula; Tutorials in Introductory Physics (TIP), Open Source Tutorials (OST), and Investigative Science Learning Environment (ISLE). We interviewed the designers, converted interviews to transcripts, and coded the transcripts. We used a constant comparative approach for the coding process (8) which consists of two parts.

1) Freely coding segments of data, inventing new codes, and comparing segments of data with one another (open coding).

2) Relating the codes to one another to develop categories (axial coding). Based on our investigation, we found a model known as revisiting cycles.

In this work we identified a strategy we labeled as revisiting. Revisiting means an activity asks students an initial question, then discusses the same question for the second time by guiding students to the initial question. We found four different approaches to achieve this strategy.

1) Asking students to check their initial answers with an experiment.

2) Asking students to work alone and then share their ideas as a group.

3) Asking students to verify their answer with the instructor/TA.

4) Providing answers to initial questions for students.

Therefore, revisiting consists of two parts: an “initial question” supported with a “revisit.” Revisiting became a category in our codebook (Table 3.3).

2.3 Thinking like a physicist and Conceptual learning

In the middle stage of the research, we published a second Physics Education Research Conference (PERC 2017) proceeding paper (15). We interviewed more curriculum designers and analyzed their activities. Moreover, We included the AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum 2014 in our analysis. This document is influential in the field of physics education and comes from the leading professional society in the field. It lays out guidelines and best practices for activities produced by a group of researchers and educators respected in the field. Again, we found revisiting cycles and argued that designers' perspectives about conceptual learning and scientific thinking influence the design features they use in their activities.

1) Thinking like a physicist

One of the AAPT recommendations is "constructing knowledge" which should help students to "thinking like a physicist." However, we noticed designers of research-based activities have different points of view on this. An ISLE designer believes that "you are thinking like a physicist in your life already. If you look at little children, they actually do ISLE all the time." Clearly, "They would come up with hypotheses to explain evidence, with multiple hypotheses, and then they would actually systematically test them without saying I'm testing a hypothesis." On the other hand, one of the University Modeling Instruction (UMI) designers displayed a different idea about "thinking like a physicist." The designer stated "We have been developing using sort of a Modeling cycle which is based on a learning cycle approach, but the idea being that the epistemological background is that we believe that science proceeds through the building, validation, and testing of models."

2) Conceptual learning

Conceptual learning is one of the main goals of many research-based activities. For some research-based activities such as Workshop physics, it is the primary goal. One of the designers of Workshop Physics stated that "these hands on [research-based activities] are directed at trying to overcome conceptions students tend to bring to a course that makes it difficult to learn." However, for ISLE and UMI, conceptual learning is a secondary goal. For instance, one of the UMI designers said that the "goal of Modeling is to use these models and in order to do that you have to ... stop by some of these other things as well. You know, conceptual reasoning, active learning, these sorts of things." Moreover, one of the designers of ISLE explained that "the main goal of ISLE is to help students learn to think like physicists, [but] we use traditional PER assessments as well."

2.4 Three clusters in physics activity design

There is substantial demand for reforming physics curricula to include more interactive engagement activities and student learning improves by using these activities (16). Activities come in many forms such as tutorials, labs, worksheets, and "ponderables" (17). Sometimes these activities are used individually by instructors in lectures or laboratories and sometimes they are integrated into curricula such as SCALE-UP (17; 18), University Modeling Instruction (19) and Workshop Physics (20; 21). Instructors often turn to research-based activities as reliable methods that have been proven to help students learn. Activities are typically described as accounted research-based when results demonstrating their effectiveness have been peer-reviewed and published. For example, RealTime Physics (RTP) (22; 23), Socratic Dialog-Inducing (SDI) Labs (24; 25; 26), Cornell Thinking Critically in Physics Labs (CL) (27; 28), and Scientific Community Laboratories (SCL) (29) are examples of research-based activities. However, there may sometimes exist a mismatch between the goals of the instructors and those embodied in the activities or instructional strategies. According to Henderson *et al.* (30) approximately one-third of instructors abandon the use of research-based instructional strategies (RBIS) after attempting one or more strategies. In other work, Henderson and colleagues (31) conclude that when faculty used RBIS " (i)n many cases they reinvented instruction that was missing important fundamental features of the intended instruction and/or conflicted with recommended practices." They report that instructors adapt and reinvent curricular materials due to "the personal nature of teaching and the unique instructional environments." In a project involving visits to ten U.S. universities teaching SCALE-UP physics courses, one of my colleagues (Joshua Von Korff) observed that all but one institution used activities created by instructors rather than research-based activities. Many of these activities were observed to be research-inspired, meaning that the materials used principles developed through education research. Rather than using research-based activities as published, instructors chose to modify them or create their own activities to meet particular goals. The process of redesigning activities may be expensive and time-consuming, so developing principles that assist instructors in creating activities that meet particular pedagogical goals could be very useful. To design research-inspired materials, one needs to know useful design principles, and the efficacy of the materials will depend entirely on the efficacy and accuracy of those principles. The intent of this publication is to help instructors design research-inspired activities by providing them with those principles (15; 13) Meltzer and Thornton reviewed many active-learning instructional methods in physics and discussed their effectiveness for student learning (32). They also identified some common characteristics among research-based active-learning instruction in physics such as student ideas are elicited and addressed, students express their reasoning explicitly, qualitative reasoning and conceptual thinking are emphasized, problems are posed in a wide variety of contexts and representations, instruction frequently incorporates use of actual physical systems in problem solving, and instruction emphasizes linking of concepts into well-organized hierarchical structures. The present work identifies common characteristics of research-based physics activities but also examines the connection of these characteristics to particular design goals. Design goals for activities reported in the literature include learning concepts (19; 21; 22; 24; 33; 34), thinking like physicist or scientists (35; 36), understanding measurement and uncertainty(27; 29), designing experiments (27; 29; 33), constructing models (19; 37), and improving students' beliefs about the nature of experimental physics (38).

Our long-term research goal is to find a practical way to help instructors who want to use the features of different research-based activities to design their own activities. The present work takes a primary step toward this goal by grouping research-based activities according to their design features. This work seeks to answer the following research questions:

1) How can we cluster the research-based activities based on their design features?

2) How are these clusters related to the goals of the designers?

Preliminary results from this project have been reported in which we investigated several uses of representations in evidence-based and non-evidence-based physics activities (39), the role of revisiting as an essential and common technique in tutorials (13), and analyzing several design philosophies revealed through interviews with designers of research-based activities (15).

CHAPTER 3

Data collection and analysis methods

3.1 Introduction

This chapter will explain the sources used in our research, the coding process, finding agreement on our codebook, and analyzing data quantitatively using k-mean cluster analysis. Most of the information in this chapter is mainly from submitted journal paper to Physical Review Physics Education Research (40).

3.2 Sources

We used three primary sources to develop the coding scheme in this work: 1) interviews with the designers of research-based activities, 2) 2014 AAPT lab report, and 3) published research-based activities. The coding system was then used to characterize research-based activities from eleven research-based curricula. To improve our qualitative research reliability and conclusions, we utilized these three data sources and diverse methods to analyze the data (triangulation approach). This research approach helped us to investigate our qualitative data from different perspectives (41; 42).

3.2.1 Interviews

We performed one-hour, semi-structured interviews with 15 designers of some of the most frequently used research-based activities, including Investigative Science Learning Environment (ISLE), University Modeling Instruction (UMI), RealTime Physics (RTP), Workshop Physics (WP), Tutorials in Introductory Physics (TIP), and Open Source Tutorials (OST).

The main focus of the interviews was to ask about the principles or techniques used to design activities, how these design principles help students achieve the goal of the activities, and what the similarities and differences are between their activities and other commonly-used activities. We aimed to learn more about the goals and purposes of designers than is revealed in published articles. Participants expect that we will not reveal their names, but they may be identifiable based on only one or a few authors for each curriculum. The interviews were video recorded, transcribed by one researcher, and another researcher checked the transcriptions for accuracy. A number of people participated in the transcription process, mostly undergrad students who selected education courses and worked as colleagues each semester. We used the interview protocol shown in Table 3.1. We designed the interview protocol questions to help us to analyze research-based activities effectively. Designers use some examples of their research-based activities in the literature to help the reader understand them better. However, these examples are general and sometimes did not address details of activities we were analyzing. Therefore, designers were asked about specific examples from activities that we had already analyzed in our research to make sure that we can relate these data together. Also, finding differences and similarities among activities helps us find more reliable codes to categorize research-based activities. We asked additional questions such as, "What are your favorite tutorials or activities (that you or others have written)?" Then we can discussed these further in the interview to gather as much information as possible about their goals and preferences. We asked in question 2 of the interview protocol for designers to define and elaborate on their design principles (DP). We might already have known of a place in the tutorial or lab activities that we think used this DP. Sometimes, we determined that the examples considered so far all have a certain property, we wanted to find out whether use of this design principle was always associated with that particular property. At other times, we had a definition of a design principle that seemed more intuitive to us than the participants' definition. In both cases, interview questions were developed to look for negative examples so that we could be more certain of our interpretation. This is the reason why we ask questions to follow up on question F from the question 2 section. The goal of question 6 is related to one of the designers' ideas that a given activity does not always work everywhere.

3.2.2 AAPT lab report 2014

In 2014 American Association of Physics Teachers (AAPT) published Recommendations for the Undergraduate Physics Laboratory Curriculum (14) suggesting specific learning goals for introductory physics laboratories such as constructing knowledge, modeling, designing experiments, developing technical and practical laboratory skills, analyzing and visualizing data, and communicating physics. This report was designed and published by the AAPT Committee on Laboratories. It is a community of PER experts designated to teaching physics as an experimental science in both conventional (in-class) and non conventional (out-of-class) environments at all levels of teaching physics. Members of the Committee on Laboratories regularly have meetings and design workshops for faculties in the different meetings to help physics instructors improve their teaching laboratory skills. This report is influential in the field and considered a reliable source in the AAPT community. Different researchers have used it as a reliable resource in their research and followed their recommendations.

	Main question	Follow up questions
1	Basic questions	**A)** Which tutorial(s) or lab(s) did you design? **B)** When was the work done (in what year?)
2	What design principles (DPs) / techniques do you know of?	**A)** What is the function/purpose of this DP? **B)** What is the definition of this DP? **C)** How can I tell by looking at an activity whether it uses this DP? **D)** Give an example of this DP (in a particular activity? Any activity?) **E)** How does this DP help students change their ideas in this particular tutorial/lab activity? **F)** Is there a negative example that is X but not this DP? That is this DP but not X? **G)** Does tutorial Y make use of this DP?
3	How is your set of tutorials/labs similar or different from another set?	**A)** How is your implementation of a common DP similar to theirs? **B)** How is your implementation of a common DP different from theirs? **C)** How / why do you use different DPs from theirs?
4	How does your tutorial/lab relate to theory? (Misconceptions, difficulties, resources, Redish's five princ iples, or anything people have already said in cognitive science or PER about learning)	**A)** How does your tutorial/lab / DP relate to any theory of the participant's choice? **B)** How does your tutorial/lab / DP relate to theory T, chosen by the interviewer? **C)** What does theory T say, in your opinion?
5	Asking about known DPs or frameworks.	**A)** Do you agree with DP or other ideas that someone else suggested? **B)** How would you define it / name it / describe its purpose?
6	Student populations.	**A)** Do these DPs work everywhere in your experience or more at some schools than others? **B)** Do these DPs work with all levels of students (calculus/algebra) or more with some than others?
7	Miscellaneous question types (to be used all the time)	**A)** Active listening: trying to repeat back exactly what the other person said, with minimal interpretation. **B)** Member checking: repeating back with some interpretation and checking if we are right.

Table 3.1 List of questions that we used in our interview protocol.

3.2.3 Research-based Activities

In this work we evaluated 66 introductory college level laboratory and classroom activities from 11 research-based curricula. Most of these were chosen based on their effectiveness as reported by Von Korff *et al.* (43) who evaluated the results of Force Concept Inventory and the Force and Motion Conceptual Evaluation for different interactive engagement teaching techniques published between 1995 and 2014. Activities from two research-based curricula were included because they were studied in earlier work by Thacker *et al.* (34). In that study FCI gains from "physics education research-informed materials" were compared to traditional activities in a large university. Two additional resources were included because they are widely-used research-based activities recommended on the PhysPort website (12). Some of the activities were free to download, and others required that we receive permission from the authors or designers. Among the activities studied here, ISLE is not only a set of activities, but also a design philosophy; so there may be many labs that are compatible with ISLE. We used a particular set of labs created by the authors of the ISLE design philosophy. The sets of all research-based activities investigated are explained as follows.

3.2.3.1 RealTime Physics (RTP)

The primary goals of RealTime Physics (RTP) are to design a series of complementary laboratory activities that improve students' conceptual understanding and quantitative laboratory skills based on PER. RTP activities are designed for non-integrated physics courses (44). RTP activities are designed for introductory level for college or university physics courses. They are invented explicitly for instructors who want to use the laboratory's active learning method without changing the introductory physics course's overall construction. Students build their physical models built on "observations, experiments, mathematical modeling, data analysis, and simulations (12)."

3.2.3.2 Investigative Science Learning Environment (ISLE)

ISLE is a philosophy of learning and teaching physics first designed by Eugenia Etkina and Alan Van Heuvelen. ISLE uses specific models to help physicists to build their knowledge. According to Etkina *et al.* (7), "Students construct physics concepts and develop science process abilities emulating the processes that physicists use to construct knowledge." The designers spell out four steps to this process.

A) Students decide to inquiry about a physical phenomenon.

B) Students collect data about the phenomenon, recognize patterns, and generate different explanations for why the phenomenon is happening.

C) "They test their explanations by conducting one or more testing experiments." The main purpose of the test experiment is to reject explanations instead of "prove" them. Ideas that pass from step 3 are "kept and re-tested" for subsequent experimentation.

D) Students find applications for their ideas to solve real-world problems. "The cycle repeats twice, first qualitatively, then quantitatively" as illustrated in Figure 3.1 (45).

3.2.3.3 Scientific Community Laboratories (SCL)

According to Lippmann, Scientific Community Laboratories' main goal is to help students practice a primary knowledge of measurement and uncertainty by design and operate their experiments to "produce, analyze, and evaluate scientific evidence." In this curriculum, students learn the "concepts underlying uncertainty in an experiment, (measurement concepts)," and they use it to design experiments and evaluate their results. "Measurement con-

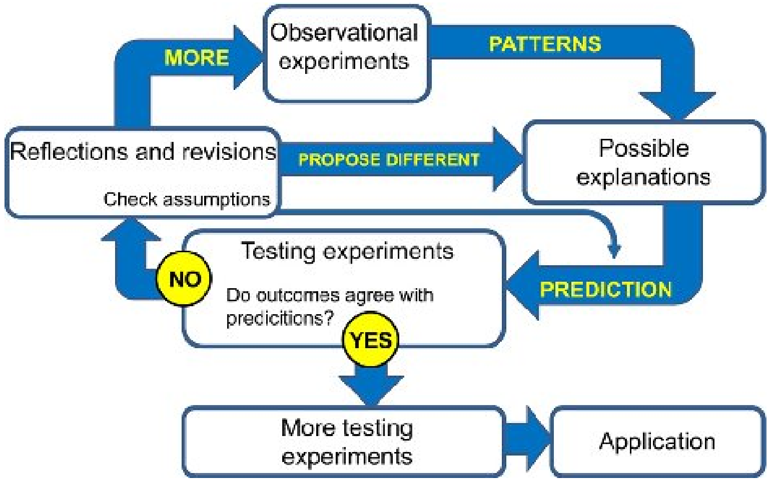

Figure 3.1 Investigative Science Learning Environment (ISLE) cycle (Taken from Ref. 43)

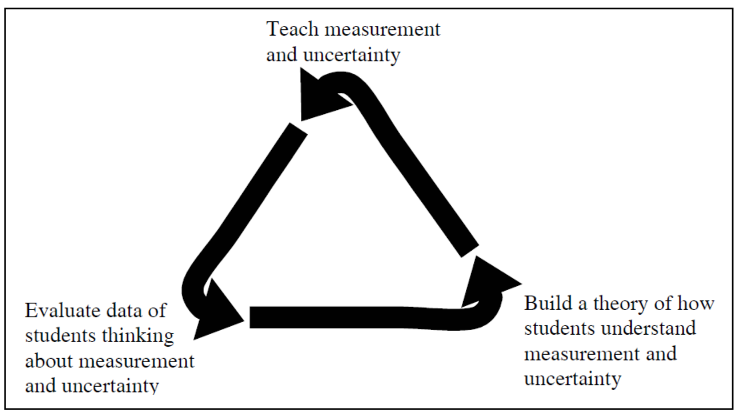

Figure 3.2 The goal of this Scientific Community Laboratories and how it works (Taken from Ref. 29)

cepts" are similar to the physics concepts such as normal force that help students construct force and motion knowledge. In the same manner, "measurement concepts" are essential understanding of measurement. (29) For achieving this goal, they analyze "students' initial state (what they already know), final state (what they need to know) and have a model of how cognitive change occurs," as illustrated in Figure 3.2 (29).

3.2.3.4 Physics Department, Texas Tech University labs (TTU)

Texas Tech University labs (TTU) are PER-informed labs designed by Texas Tech University instructors using "PER literature, other PER-based instructional materials, and pedagogical content knowledge." The design goals of TTU are to address common student misconceptions by using elicit, confront, and resolve strategy by examination and conversation with other students and the TAs. It is the same strategy used by Tutorials in Introductory Physics. TTU also concentrate on "quantitative measurements and observations, taking data, graphing, analyzing, and interpreting it." (34)

3.2.3.5 The University of Illinois labs (UI)

The University of Illinois labs (UI) are PER-informed labs developed at the Department of Physics of the University of Illinois. UI are designed to improve University of Illinois introductory physics courses by inspiring RealTime Physics. They address common misconceptions of students by using an active engagement strategy.(34)

3.2.3.6 Tutorials in Introductory Physics (TIP)

The Physics Education Group at the University of Washington designs tutorials in Introductory Physics (TIP). It is a guided inquiry worksheet design for teaching physics to small groups of students in the recitation section of intro calculus-based physics. Instructors involve group discussion by using Socratic dialogue. Worksheets use the "Elicit-Confront-Resolve" method. It means each question elicits students' misconceptions and guides them in specific procedures to confront and resolve their misconceptions. (12; 2) TIP is supported by a rigorous research and redesign process, and different works of literature prove that students can learn from it. TIP can be practical even though instructors are not familiar with the PER process behind it. (12) Some physics education researchers believe that the "elicit-confront-resolve" strategy used in TIP "may give students a sense that their intuition about physics is always wrong and leads to decreased self-efficacy." (12)

3.2.3.7 Socratic Dialog Inducing Laboratories (SDI)

Richard Hake designs Socratic dialogue-inducing (SDI) labs to help students investigate "Socratic questions on physics, science, and ways of thinking, culminating." (24). They use "interactive engagement" techniques and are intended to improve students scientific skills such as "appreciate the need for operational definitions; use and interpret pictorial, graphical, vectorial, mathematical, and written representations; and consider dimensions, thought experiments, and limiting conditions." Different researches prove that the SDI lab helps students to think like a scientist. (24)

3.2.3.8 Workshop Physics (WP)

Workshop Physics is a method of teaching calculus-based introductory physics that is designed for interactive classes. It emphasizes students' collaboration and observations. Standard WP courses consist of three two-hour-long sessions each week, and students practice Activity Guides. In WP, enhancing students' scientific inquiry skills are more significant than problem-solving. They teach students scientific skills by emphasizing "observing phenomena, analyzing data, and developing verbal and mathematical models to explain observations."

Designers of WP believe students can teach physics to each other better than instructors. The instructor's role is to design the learning atmosphere, guide investigations, and participate in Socratic dialogue with students.(46)

3.2.3.9 University Modeling Instruction (UMI)

University Modeling Instruction concentrated on the concept that "physicists reason from mental constructs known as models." UMI creates model building by using various representations to describe particular physical circumstances. By repeating the process of using different representations and analyses, students recognize common features and patterns and then interpret and coordinate them into a "general model" that fits a comprehensive level of circumstances (19).

Following are features of UMI activities:

A) Substantial importance on exploration in each activity. Activities tend not to tell students what they should do in each step but instead help them 'muck about.'

B) Students expect confirmatory labs and expected to get the correct answers. However, it contrasts UMI activities' goals since they should look for "patterns within phenomena."

C) Because finding patterns among data is a challenge, so students have different options for writing their activity reports. For assisting students, instructors encourage them to use two options. First, when activities are "heavily scaffolded," the instructor motivates them to "complete the activity and turn in a completed worksheet that is then graded for completeness." Second, instructors ask students to explain the pattern(s) they discovered and provide evidence to support their claim(s) (47).

3.2.3.10 Open Source Tutorials (OST)

Open Source Tutorials (OST) are guided-inquiry activities for small students' groups in the tutorial section of intro algebra-based physics courses and may not work with other courses. Instructors join students' groups in Socratic dialogue. Tutorials "refine students' productive intuitions" and improve metacognitive thinking (12). There are plenty of resources on the OST websites to help instructors modify tutorials and train TAs, including "interactive lecture demonstrations, test questions, instructor's guides, and TA training workshops." (48) These activities are designed to use as supplementary for lectures. They concentrate on developing "students' productive resources" rather than "confronting misconceptions." The Physic Education research behind OST is not as powerful as for Tutorials in Introductory Physics (12).

3.2.3.11 Cornell Thinking Critically in Physics Labs (CL) Version 2018

Holmes *et al.* (28) design a learning framework that uses specific decision-making cycles to help students quantitatively compare data between data sets and models. This cycle of learning can improve students' critical thinking about scientific evidence. This structure is suitable for any "data-driven science-learning setting" with chances to improve the data or models. They mention that this structure can improve students' critical thinking as illustrated in Figure 3.3 (28).

3.3 Analysis

Figure 3.4 shows a diagram representing our coding process.

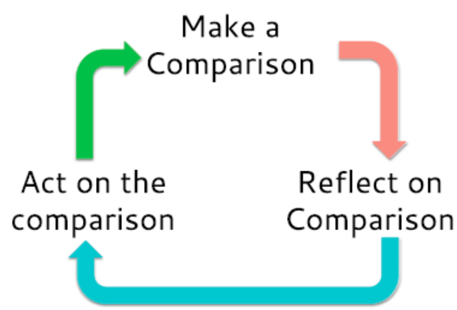

Figure 3.3 The goal of this Scientific Community Laboratories and how it works (Taking from reference 28)

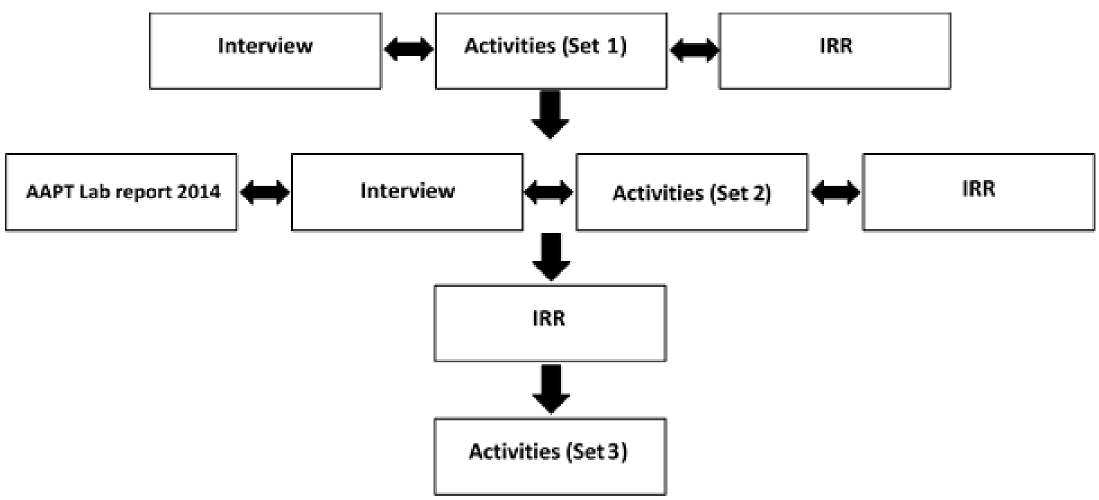

Figure 3.4 Four steps of the coding process. Step 1: coding of interviews and 25 different activities and evaluating by IRR. Step 2: coding of AAPT lab report 2014 and 40 additional randomly-chosen activities, recoding of interviews, and evaluating by IRR. Step 3: randomly choosing one activity from each research-based curriculum and evaluating by IRR. Step 4: coding of five additional randomly chosen activities from each curriculum.

Since activities by their nature ask the students to perform tasks which are intended to help them learn, we coded the designer interviews for expected student actions in order to understand the goals of the activities. We used the constant comparative method to develop the coding scheme and we met regularly to discuss and debate the codes and their arrangement into categories (axial coding) (8). We used this scheme to code 25 different activities (referred to as Set 1), mainly from the same research-based curricula as the interviewees. At each step the coding scheme was validated for consistency using Cohen's Kappa to evaluate inter-rater reliability (IRR) (49). For the IRR process, we count codes one time if we observed them in the minimum possible unit for coding. For example, if this minimum possible unit is a paragraph and we observe a code three times, we count it only once. The reason behind this policy is that consecutive features of the same type within a paragraph were generally closely related. After finishing the IRR process, the researchers discussed the results, then eliminated or combined codes and rewrote code definitions as needed. Memos helped us to write our thought process and refer to them in next steps (6). Preliminary results from this work were published in 2016 (13). The code list from this first stage was used to analyze the introductory level recommendations in the AAPT lab report 2014 and 40 additional activities chosen from all eleven research-based curricula (Set 2) and to reanalyze the designer interviews. Additional results were reported by the authors at this stage of the investigation (15). Again, the researchers discussed and revised the code list and categories informed by a determination of IRR. The code list and categories were then evaluated by two independent physics education researchers to provide feedback. Small changes in the categories and their constituents were made in response to this feedback.

The final scheme consisted of 49 codes in ten categories as shown in Table 3.3. One activity was randomly chosen from each research-based curricula for coding and evaluation of IRR. This evaluation yielded an average Cohen's kappa of 0.78 per activity. According to Everitt (50) this value of Cohen's Kappa is “satisfactory or solid agreements”, and according

List of research-based curricula	Cohen's kappa
Cornell Intro labs Mechanics 5 Version 2018	0.653
Open Source Tutorial 2: Backward acceleration	0.775
University Modeling Instruction week 14	0.782
Workshop Physics: Unit 6: Gravity and projectile motion	0.925
Socratic Dialogue-Inducing Lab 4	0.762
Tutorials in Introductory Physics: Wave properties of light	0.677
Texas Tech University lab: Mechanics 14	0.865
University of Illinois labs 8	0.758
RealTime Physics Lab10: Electromagnetism	0.874
ISLE Lab 9: Refraction	0.760
Scientific Community Laboratories: Damped Oscillations$_{p}artI_{a}ndII$	0.707

Table 3.2 List of 11 research-based curricula whose activities were investigated for IRR and their values of Cohen's kappa.

to Fleiss *et al.* (51) it shows "excellent agreement". The minimum and maximum values of Cohen's kappa we obtained were 0.653 and 0.925, respectively, as seen in Table 3.2.

After achieving good IRR results, one of the researchers randomly selected five additional research-based activities from each research-based curricula for coding. The coding results from six activities from each of the eleven research-based curricula were then used for the cluster analysis.

3.4 k-means cluster

K-means analysis clusters data by "minimizing Euclidean distances" among them in the multi-dimensional space represented by the data. k-means cluster analysis follows the following steps:

1) The researcher selects the number of clusters (for example, two clusters). 2) An algorithm assigns initial center locations to each cluster randomly.

3) Each data point is then assigned to the cluster whose center it is nearest.

4) A new center location is calculated for each cluster as the centroid means of its clus-

Category	Code	Definition
Observation: The process of collecting data by observation and using it	Observe Data from Equipment	Refers to a physics or math question or instruction about observation and recording of data from equipment by students.
	Observe Data from a Simulation	Refers to a physics or math question or instruction about observation and recording of data from simulation by students.
	Use Observed Data from Equipment	Refers to a physics or math question about data previously observed and recorded from equipment by students.
	Use Observed Data from Simulation	Refers to a physics or math question about data previously observed and recorded from simulation by students.
	Extract Data from Video	Asks students to extract data from videos during the observation.
Prediction: The process of prediction an experiment or making hypothesis by students and comparing the result of the experiment with the prediction	Make Prediction	Asks students to make a prediction, which means that (1) an experiment will be done in the future and (2) the students are asked to figure out the result before experimenting.
	Check Prediction	Asks students to decide if a prediction was consistent with their observation.
Spoken Representation: Communication among students, instructor, and class	Check with TA or Instructor	Asks students to talk to the instructor about some work they have been doing.
	Group discussion	Asks students to talk to their group.
	Class discussion	Asks students to talk to the whole classroom.
	Symposium	Asks students to visit or talk to other groups.
	Show Whiteboard	Asks students to show their whiteboard to an instructor, another group, or the whole class.
	Think/Pair/Share	Refers to thinking individually, comparing answers with other group members and resolving any conflicts.
Design: The process of design, improvement, making hypothesis for an experiment or math	Procedure Design	Designing procedure, could include describing an experimental procedure invented by the students, explaining how the students invented an experimental process, or explaining what decisions the students had to make to invent the experimental procedure.
	Improve design	Asks students to improve their previous designs.
	Choose Question to Investigate	Asks students to choose an open-ended inquiry question.

procedure by students	Designing Math Procedure	Asks students to design/state/invent/improve a mathematical or quantitative procedure they will use before they use it.
	Making Hypothesis	Asks students to make a hypothesis that they have devised.
Qualifications: Asking students for their assumptions, simplifications, limits, error analysis and reasonableness of their answers	Assumptions, Simplifications, Limits	Asks students about their assumptions, simplifications, and limitations with their model or way of understanding a physical situation
	Error and Uncertainty	Asks that the students either give a qualitative discussion of error or estimate or quantify the uncertainty or error.
	Reasonableness	Asks about the reasonableness of results or answers. Also, includes questions about the nature of reasonableness, or what it means for something to be reasonable
Miscellaneous: Any code not in another category	Most Important Concept	Asks students what the most important concepts are
	Goal or Purpose	Asks about what students will have or be able to do or what question the students will answer by the end of the lab.
	Non-observation Question	Students answer physics or math questions that do not use data from a previous measurement or observation and are not a prediction question.
	Instructor Guide	Tell the instructor what to do (as opposed to telling students what to do). The activity might tell the instructor to lecture in a certain way, to help students' groups, or to give a demonstration
	Real-world Example	Asks students for a real-world example.
	Computer Data Analysis	Asks students to use a computer to analyze existing data numerically. The computer processes and displays numbers or equations such as errors, means, or the parameters of a curve fit.
	Grading Rubric	Gives a grading rubric to students showing how students' work will be assessed.
	Generalization	Asks students to Identify trends or reason by induction to produce generalizations.
	Calibration	Students are instructed to calibrate equipment, such as a scale.
	Ethics	The instruction mentions ethical considerations, such as plagiarism.
	Notebook	Students should write something in a lab notebook, report, or other documents that are separate from the activities' questions. That means students have to organize their responses themselves.
Epistemology:	Epistemological Question	The activity asks general questions about how to think, how to learn, how to proceed with certain kinds of physics problems, or how to go about doing physics.

Refers to epistemological questions		
Written Representation: Different kind of written representations	Written Word	Asks students to write their idea or explain something.
	Math	Asks students to write variables, numbers, equations, and units.
	Student Chosen Representation	Asks students to produce a model, choose one or more representations, or to use multiple representations, but the specific representations are not named. The students are not told what to do with representations.
	Diagram	Asks students to draw a diagram.
	Graph	Asks students to draw graphs.
	Multiple Choice	Asks students to choose from several answers (could be given as words, pictures, or equations) or answer "yes/no" questions (or questions that implicitly only have two answers).
	Ranking Task	Asks students to rank items (e.g. from most to least).
	Pie Chart	Asks students to draw a pie chart.
	Bar Chart	Asks students to draw a bar chart.
Evidence	Evidence	Asks students to use data to support a claim. Students should produce both claims and evidence.
Revisiting: Asks an initial question, then addresses the same question a second time. Does not include predictions.	Revisit with Reasoning	Any revisiting pattern that requires students' reasoning and doesn't fit the other strategies. A common wording for this strategy would be: "is your answer consistent with …".
	Procedural	Asks any revisiting question that uses traditional style procedures, such as plugging numbers into a formula.
	Check Printed Document	Asks students to check their answers against a printed document, such as a photograph or table provided by instructor. The word "consistency" may be used.
	Statement to Agree or Disagree	Statement to agree or disagree with (or in what way do you agree or disagree). Involves one or more statements often attributed to fictitious students for students to agree or disagree.
	Telling Answer	The activity or instructor tells students the answer to the question (or a set of questions) after the students answer the question.
	Revisit by Video File	Student checks their answer against a video file.

Table 3.3 List of the all the codes and categories with the six features used in the k-means cluster analysis highlighted.

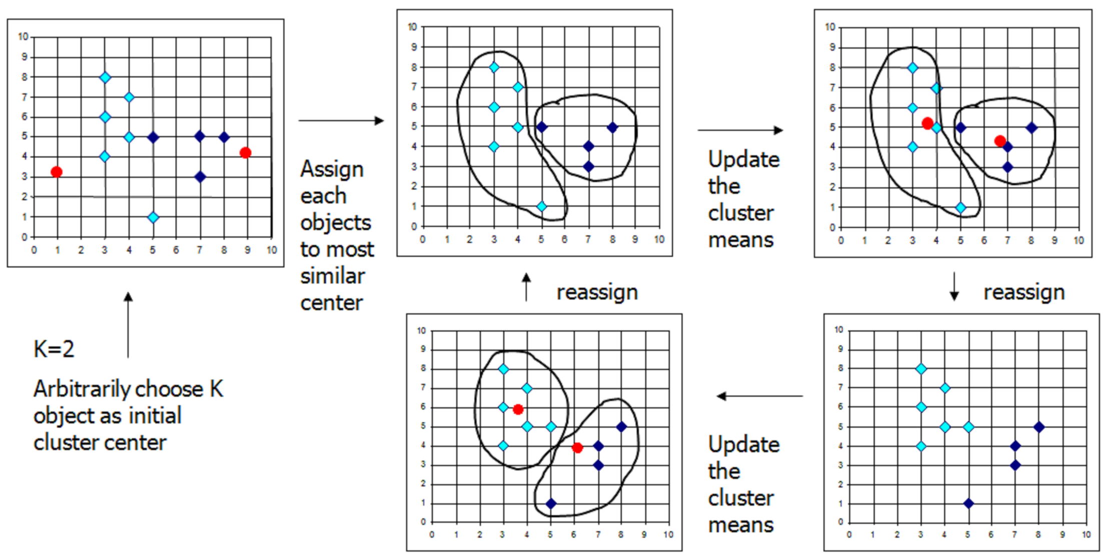

Figure 3.5 k-means cluster analysis for two clusters (Taken from Ref. 51).

tering variables.

5) Steps 3 and 4 are then repeated many times, assigning data points to the cluster with the nearest center and then recalculating center locations, until no further change occurs.

The algorithm reaches its optimal situation, and the final cluster achieves its solution (52) as illustrated in Figure 3.5 (53).

The frequency at which codes appeared varied greatly among the activities and research-based curricula we analyzed. We applied a k-means cluster analysis (54) to group the activities according to the pattern of codes. According to Formann (55), applying a k-means cluster analysis to data with m features requires a minimum of 2^m data instances. Since we analyzed a total of 66 research-based activities with our final coding scheme, we were limited to choosing six features for the cluster analysis. Of the 49 codes in ten categories, we identified one code (Non-observation Questions) and five categories (Observations, Prediction, Spoken Representation, Design, Qualifications) as the best features for the cluster analysis (highlighted in Table 3.2). Five categories appeared in our analysis with high frequency and also showed significant variations among the activities. For example, the Written Representation category had high frequency but showed little variation among the activities since this is an extremely common expectation for student action. In contrast, the Design category appeared with high frequency in some activities and rarely in many others. The only single code selected as a feature for the k-means cluster analysis, Non-observation Questions, was coded when students were asked a physics or math question that did not use data from a previous measurement or observation and was not a prediction of future observations.

The k-means cluster analysis was performed in Python (Jupyter Notebook version 5.7.8) using the KMeans function in the sklearn.cluster package. Each of the 66 activities were points in the analysis and the Euclidean distance was used to measure similarities among points (56). We define v_{ij} to be the value of the frequency for the i^{th} feature in the j^{th} activity. For instance, if an activity had three of code X and one of code Y and no other codes, their frequencies would be 0.75 and 0.25, respectively. Finally, because some student actions are more prevalent than others, we normalize each feature's frequency over all activities using z-scores to determine the final values of v_{ij}. These frequencies locate each of the 66 activities in a six-dimensional space. The goal of k-means cluster analysis is to locate the N cluster centers that minimize how far activities are from their cluster center. The center of each cluster C is defined as

$$v_{iC} = \frac{1}{N_C} \sum_{j \in C} v_{ij}$$

where the sum is taken over all activities in the cluster, and N_C is the number of activities in the cluster. Each activity is taken to be a part of the cluster whose center is closest to that activity using the Euclidean distance,

$$D_j = \sqrt{\sum_{i=1}^{6} (v_{ij} - v_{iC})^2},$$

between them in a six-dimensional space. The clustering process begins by selecting N initial centers at random and determining the activities that are nearest those centers. After each activity is assigned to a cluster, a new center for each cluster is computed. Since this may cause some activities to now be closer to the center of a different cluster, the distances D_j are recalculated and each activity is again assigned to a cluster and a new cluster center computed. This process repeats until there is no longer any change in the cluster assignments.

However, this result could be a local optimum solution rather than a global optimum. Therefore, the process is repeated 1000 times with new randomly chosen initial centers. For each repetition the quality of the clustering is evaluated by calculating the inertia of each clustering solution, $I = \sum_j D_j^2$. The cluster arrangement with the lowest inertia is taken as the final solution for each value of N. Since the best choice for the number of clusters is not known, the "elbow method" (57) was used. To apply the elbow method, we computed the inertia I_N for each number of N clusters, with N between 1 and 5. The inertia is a measure of the quality of fit of the clustering with a larger inertia meaning a worse fit. As such, I_N should decrease when new clusters are added and N is increased. As illustrated in Figure 3.6, the inertia gradually decreased as the number of clusters increased from one to five as

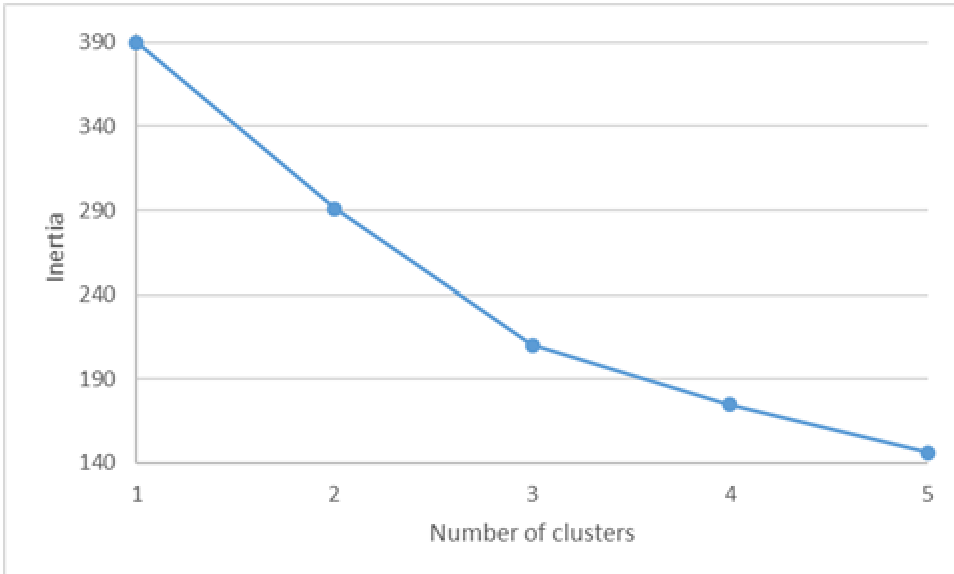

Figure 3.6 Graph of smallest inertia (arbitrary units) achieved for each number of clusters. Smaller inertia corresponds to more compact clustering. As expected, the inertia decreases whenever the number of clusters is increased.

expected. The elbow method determines the relative improvement of the fit by the addition of the Nth cluster by maximizing $(I_{N-1} - I_N) - (I_N - I_{N+1})$. When this value is large, it means that N clusters produce a much bigger improvement over $N-1$ clusters than $N+1$ produces over N clusters, suggesting N clusters as the optimal choice. Figure 3.7 shows the application of the elbow method to our data and reveals that the largest improvement is achieved when adding the third cluster.

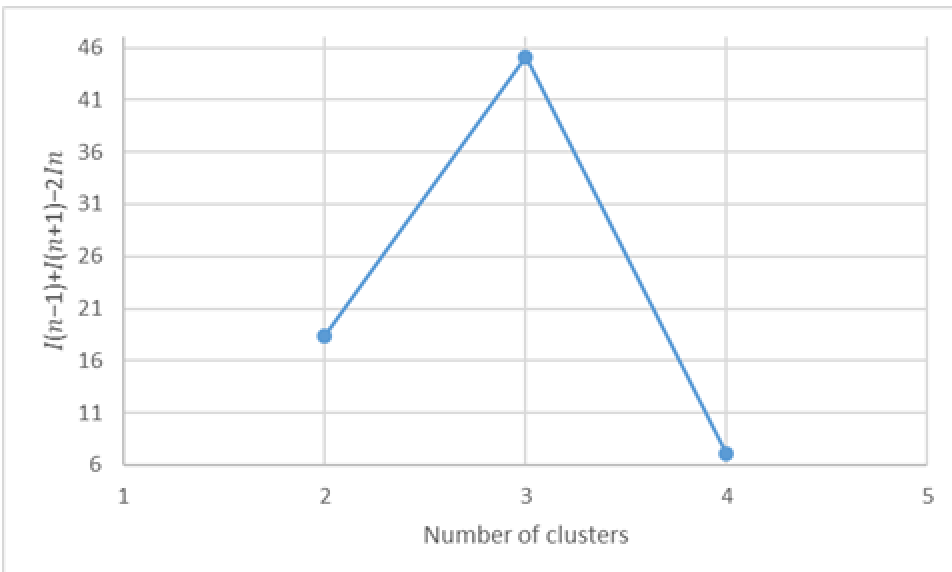

Figure 3.7 Graph of decrease in inertia when increasing from N-1 to N clusters minus decrease in inertia when increasing from N to N+1 clusters (arbitrary units) vs. number of clusters. Elbow method analysis exhibits the highest value when increasing the cluster number has the largest relative impact which occurs for N=3 in our data.

CHAPTER 4

Results and discussion

4.1 Introduction

In this chapter, I will discuss the result of k-means cluster analysis and use interviews and literature to support the results. Also, I will explain if a designer mentions specific goals in their literature, then it does not mean that these goals should be distributed equally in each activity. According to designers' experience and circumstances, they may change the weight of goals in their activities. Moreover, I will talk about the limitations of my research. Most of the information in this chapter is mainly from submitted journal paper to Physical Review Physics Education Research (40).

4.2 Results

To determine the design goals exhibited by each of the three clusters, we examined each cluster for patterns in the frequency of coding features. We evaluated those patterns by comparing them with design goals expressed in the literature and interviews. Figure 4.1 shows the average z-score of normalized frequencies for each feature for the three clusters. This analysis led us to name the three clusters Thinking like a Scientist, Learning Concepts, and Building Models, as described in more detail below. While we placed each activity in one of the clusters, all activities will do each of those to some extent, of course. But the clusters reveal that individual activities embody one goal more than the others and therefore lean more heavily on particular student actions to accomplish the goal. Statements from the

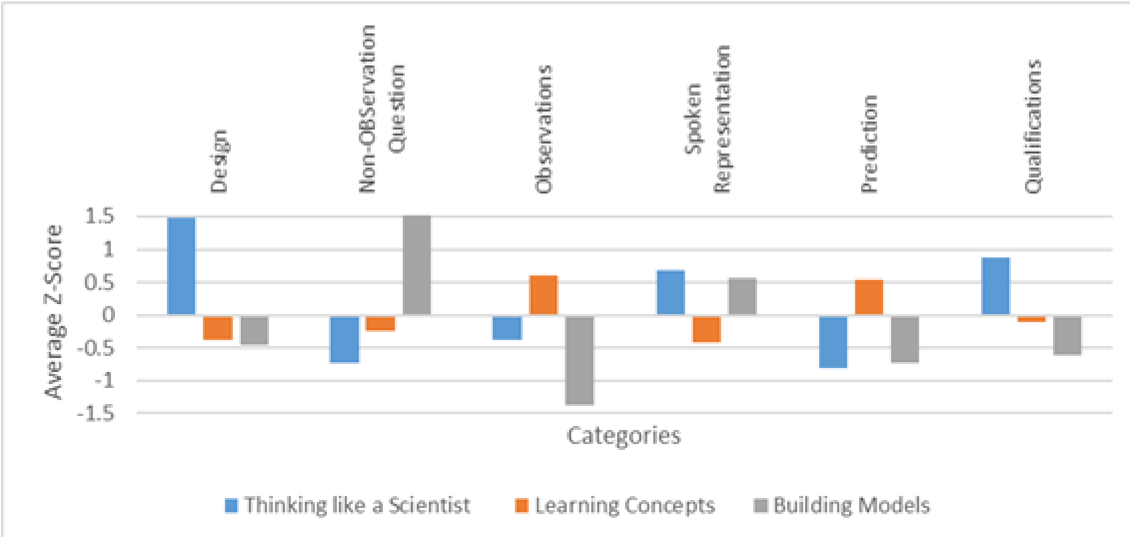

Figure 4.1 Average Z-Score of each cluster for each of the analyzed features (Design, Non-observation Question, Observation, Spoken Representation, Prediction, and Qualifications).

activity designers, in literature or interviews, also reveal these goals.

4.2.1 Thinking like a Scientist

The most frequent features observed in this cluster were Design, Produce Spoken Representation, and Qualifications as illustrated in Figure 4.1. These features show an emphasis on students performing scientific practices such as experiment design, reaching decisions by collaboration, and examining results and processes for accuracy (such as error analysis, checking assumptions and simplifications, and evaluating reasonableness). Activities in this cluster include all six CL, five SCL, and three ISLE labs as shown in Table 4.1, Table 4.2, Table 4.3, and Table 4.6.

In recent years there has been an increasing focus on explicitly teaching scientific thinking in physics courses. Holmes *et al.* (28) argue that students need to develop quantitative critical thinking and that developing this requires "repeated practice in making decisions based on data, with feedback on those decisions." Etkina and Planinšic (36) explain that the "ability to think like a scientist while solving complex problems is . . . vital ." They state that students need to be able to "formulate a problem; collect and analyse data; . . . identify patterns., . . . test ideas; . . . evaluate assumptions and solutions; ... distinguish evidence from inference; . . . argue scientifically."

According to Lippmann, the main goal of SCL is to teach "skills and techniques for creating, transforming, and evaluating scientific knowledge" (29) and that students "understand the concepts underlying uncertainty in an experiment (called measurement concepts) and be able to use that knowledge to design an experiment and interpret their data." Holmes *et al.* state that one of the goals of the CL is "thinking like a physicist" where students gain scientific skills to apply data to "evaluate models, explanations, and methods" (35). Etkina *et al.* report that designing of an experiment by students is one of the critical components of the ISLE philosophy (33). They also state that students in the ISLE classroom engage in the process's scientists use to achieve knowledge by collaborating in groups and sharing ideas. According to our interviews the designers of ISLE regard Thinking like a Scientist as an important goal of ISLE (15). Holmes and Wieman previously pointed out that ISLE and CL both focus on making decisions in the experimentation process asking students to evaluate their outcomes (35).

The most distinct feature of the Thinking Like a Scientist cluster compared with the other two clusters is the prevalence of the Design and Qualifications features. These two

List of research-based curricula	Thinking like a Scientist	Building Models	Learning Concepts
Cornell Labs (CL) Version 2018, received in person	6	0	0
Open Source Tutorials (OST) (58)	0	5	1
University Modeling Instruction (59) (UMI)	0	4	2
Workshop Physics (WP) (60)	0	0	6
Socratic Dialog Inducing Laboratories (SDI) (25)	0	1	5
Tutorials in Introductory Physics (TIP) (61)	0	3	3
Physics Department, Texas Tech University (TTU), received in person	0	0	6
The University of Illinois labs (UI), received in person	0	0	6
RealTime Physics (RTP) (62)	0	0	6
Investigative Science Learning Environment (ISLE) (63)	3	0	3
Scientific Community Laboratories (SCL) (64)	5	0	1

Table 4.1 List of 11 research-based curricula whose activities were investigated and numbers of activities in each cluster from each research-based curricula.

features emphasize student decision-making in the creation, modification, and evaluation of experimental methods. The interviews and publications from designers make clear that their goal is for students to develop the process skills used in a scientific approach. Although Spoken Representations is prominent in both the Thinking Like a Scientist and Building Models clusters, the reasons appear to differ. Designers of activities in the Thinking Like a Scientist cluster talk about scientific arguments and explanations and emphasize students' critical evaluation of their own and each other's ideas.

4.2.2 Learning Concepts

Compared with the other clusters, the most frequent features observed were Prediction and Observation as illustrated in Figure 4.1. The Learning Concepts cluster is the largest of the three clusters and includes all six analyzed activities from RTP, WP, Physics Department, Texas Tech University (TTU), and The University of Illinois labs (UI). This cluster also includes five of the SDI Labs, three activities each from ISLE and TIP, two from University Modeling Instruction, and one from each of SCL and OST as shown in Table 4.1, Table 4.2, Table 4.4, and Table 4.6.

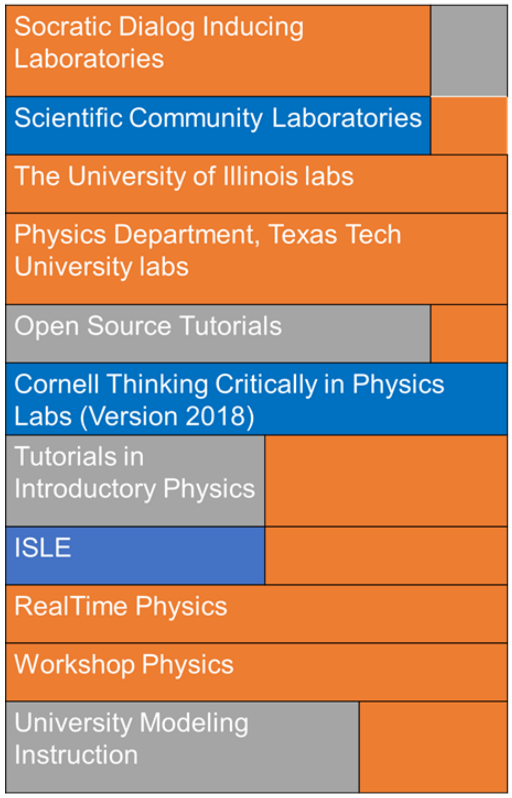

Table 4.2 Qualitative distribution of research-based activities in each cluster.Blue color: Thinking like a Scientist, Grey color: Building Models cluster, and Orange color: Learning Concepts cluster.

Cornell Intro labs Mechanics 2	Scientific Community Laboratories: Ohmic Materials part I and II
Cornell Intro labs Mechanics 3	Scientific Community Laboratories: Grandfather Clock part I and II
Cornell Intro labs Mechanics 5	Scientific Community Laboratories: Light Refraction part I and II
Cornell Intro labs Electricity 1	Scientific Community Laboratories: Magnetic Force part I and II
Cornell Intro labs Electricity 3	Scientific Community Laboratories: Double-Slit Interference part I and II
Cornell Intro labs Mechanics 4	
ISLE Lab 9: Refraction	
ISLE Lab 2: One-dimensional kinematics	
ISLE Lab Lab 9: Gases	

Table 4.3 List of research-based activities in Thinking Like a Scientist cluster.

University of Illinois labs 1	Scientific Community Laboratories: Damped Oscillations_part I_and II
University of Illinois labs 2	Texas Tech University lab: E&M 1
University of Illinois labs 4	Texas Tech University lab: E&M 5
University of Illinois labs 5	Texas Tech University lab: E&M 6
University of Illinois labs 7	Texas Tech University lab: Mechanics 12
University of Illinois labs 8	Texas Tech University lab: Mechanics 13
University Modeling Instruction week 6	Texas Tech University lab: Mechanics 14
University Modeling Instruction week 8	Tutorials in Introductory Physics: Dynamics of rigid bodies
Open Source Tutorial 2: Backward acceleration	Tutorials in Introductory Physics: Plane Mirrors
RealTime Physics Lab 3: Electric and Gravitational potential	Workshop Physics: Unit 1: Introduction and computing
RealTime Physics Lab4: Batteries, Bulbs, and current	Workshop Physics: Unit 3: One- Dimensional motion I
RealTime Physics Lab8: Introduction to capacitors and RC Circuits	Workshop Physics: Unit 5: One- Dimensional forces, Mass, and motion
RealTime Physics Lab 9: Newton's third law and conservation of momentum	Workshop Physics: Unit 6: Gravity and projectile motion
RealTime Physics Lab10: Electromagnetism	Workshop Physics: Unit 11: Energy conservation
RealTime Physics Lab 10: Two-Dimensional motion	Workshop Physics: Unit 12: Rotational motion
Socratic Dialogue-Inducing Lab 01	ISLE Lab 2: Electric Interactions II
Socratic Dialogue-Inducing Lab 02	ISLE Lab 8: Reflection and Mirrors
Socratic Dialogue-Inducing Lab 2	ISLE Lab 10: Applications of Lenses
Socratic Dialogue-Inducing Lab 4	
Socratic Dialogue-Inducing Lab 5	
Tutorials in Introductory Physics: Charge	

Table 4.4 List of research-based activities in Learning Concepts cluster.

Sokoloff *et al.* report that two purposes of RTP are to support students to “acquire an understanding of a set of related physics concepts” and to “master topics covered in lectures and readings using a combination of conceptual activities and quantitative experiments” (23). According to our interview with one of the RTP designers, they achieve this goal by using a learning cycle of prediction, observation, and comparison. Interviews also revealed that this learning cycle was used by WP. According to Laws, learning concepts is one of the main goals of WP to help students to succeed in physics, engineering, and sciences (65). According to Thacker *et al.*, teaching concepts is an important factor in PER-informed labs such as those used at TTU (34). They report that these labs are designed to "to address common student difficulties and conceptions by posing appropriate questions to elicit, confront, and resolve the difficulties." They also report that these labs let students "make observations that might challenge or contradict their present conceptual understanding and allow them to reshape their conceptual understanding through thought and discussion." Thacker *et al.* also state that "UI were designed as part of the reform of their introductory courses and were designed as an adaptation of the approach of Real Time Physics, designed to address common misconceptions through active engagement of the students in the learning process." According to Hake, the primary goal of SDI is "to promote students' mental construction of concepts" (26).

Learning concepts is one of the main goals of TIP according to Kryjevskaia *et al.* (66) who report that the "overarching goal of the tutorials is to promote functional understanding of concepts that are challenging for many students even after traditional instruction". Interviews with the designers of TIP revealed more details of the design approach and their use of Elicit-Confront-Resolve as a strategy for learning concepts. A designer explained the role of predictions and observations in student learning as a way for students to see a "confrontation between the way they were thinking and the prediction that would lead logically from that model, and yet the experiment – nature, disagrees."

Etkina and coworkers give constructing physics concepts as one of the main goals of the ISLE (67) and explain the role of prediction and experimentation in their learning cycle

(33; 67; 68). They explain that the ISLE process starts with students observing an initial experiment, then after constructing explanations they test their model with predictions and further experiments. Students may then modify and/or abandon their explanations and perform additional experiments. In interviews the designers of ISLE explain that one "can think of observational experiments as concept building experiments, ... testing experiments are concept testing experiments, you need to test it, and application experiments are multiple concepts that you have tested."

The goals expressed by the designers of activities in the Learning Concepts cluster appear consistent with the key tenets of conceptual change theory. The most frequent features show an emphasis toward students expressing their conceptual understanding and collecting data to test their ideas. González-Espada *et al.* (69) report that prediction and comparing the result of prediction with observation helps students change their conceptual understanding. According to Chi (70), using prediction and testing allows students to successfully modify their mental model. Khourey-Bowers (71) states that using predictions and hypothesis generation is one of ten strategies for conceptual change instruction which can "awaken curiosity and inspire questioning." Hesse (72) claims that an important step in conceptual change is challenging conceptions in which students predict according to non-scientific concepts followed by a demonstration event and explanation of the correct answer. So, it's not surprising that the key features of this cluster, prediction and observation, are those which bring out students' pre-conceptions, require comparison with results of experiments, and confirm or refute their understanding. While the goal of these activities may be the construction of

University Modeling Instruction week 3	Open Source Tutorial 1: Common sense and equations: Torque
University Modeling Instruction week 9	Open Source Tutorial 4: Counterintuitive idea
University Modeling Instruction week 12	Open Source Tutorial 5: What's the purpose of FBD
University Modeling Instruction week 14	Socratic Dialogue-Inducing Lab 7
Open Source Tutorial 6: Relating equations to common	Tutorials in Introductory Physics: Pressure in a liquid
Open Source Tutorial 8: Spreading forces over surfaces: Pressure	Tutorials in Introductory Physics: Wave properties of light
	Tutorials in Introductory Physics: Magnification

Table 4.5 List of research-based activities in Building Models cluster.

new mental models, the approach differs from the cluster we have labeled Building Models in that it relies on physical experimentation and observation. This is consistent with the idea of creating dissatisfaction through a "discrepant event" in conceptual change theory (69).

4.2.3 Building Models

Compared with the other clusters, the most frequent features observed in this cluster were Spoken Representation and Non-observation Questions as illustrated in Figure 4.1. Non-observation Questions are physics or math questions that do not use data from a previous measurement or observation and are not prediction questions. They tend to engage students in problem-solving, refining their intuitions, using and interpreting representations, and model building, e.g. students try to prove a formula or make a hypothesis. Activities in this cluster include five OST, four UMI, three of the TIP, and one SDI lab as shown in Table 4.1, Table 4.2, Table 4.5, and Table 4.6.

According to David Hestenes (73) "models in physics are mathematical models, which is to say that physical properties are represented by quantitative variables in the models."

Lising *et al.* (74) report that a goal of OST is student model building. According to interviews, OST activities are designed with explicit attention to the metacognitive and epistemological aspects of student learning. One aspect of this is students' revision of incorrect answers by a process of refining intuition and reconciliation. The designers explained that some OST activities use a lab without predictions since the goal of these activities is not doing experiments but instead to help students find a pattern and use more mathematical reasoning. One designer described the OST activities as having a "sense-making" feature. One of the designers of OST explained in an interview that students' spoken representations were important in the model building process. A designer states that as part of this epistemological process students are required to talk to a TA or instructor at particular points in the activity because they "wanted to provide opportunities for students to think [about] their thinking [and] instructors to engage students in those conversations and make that explicitly a part of the exercise."

According to the interview with designers of UMI, modeling is the process of building, testing, validating, and revising models. The purpose of labs is "not to confirm something that we have introduced theoretically, it is instead to introduce a new phenomenon." So, they state that "labs often are very conceptual and oriented around introducing something and bringing about a change in the modeling cycle." Brewe (19) reports that problem solving in modeling instruction differs from traditional problem solving since it is about "the application and adaptation of models." According to interviews with designers of UMI, "models are built up of representations" including spoken representations. In UMI activities, students perform "white board discussions" where students share their individual models and modify them.

TIP activities coded in this work were divided between the Learning Concepts and Building Models clusters. While some TIP activities ask for predictions followed by small experiments, according to interviews with designers some activities do not require an experiment but instead ask questions aimed only at having students construct a model. One of the designers describes their idea of model building as "breaking something up into constituent pieces and sometimes it's sort of representing sort of a complex thing." A designer explained about the goal of building models in TIP as "they (students) want to have a sort of procedure that they can say, how can I build a prediction based on think(ing) of these wave as if they are like there's this fictional pulse or, how can I predict what an extended light source, what kind of image an extended light source is going to produce based on thinking of it as many tiny sources. Or, how can I think of a circuit if I think of it as something flowing and there's pathways and barriers to that flow." According to interviews, spoken representation is one of the design principles of TIP where students are required to have discussions about their ideas in groups and at points to check with the instructor to make sure they resolve their inconsistencies. One designer explained that TIP questions are meant to be difficult for a student to answer alone which encourages students to participate in the group discussion. A key feature of the Building Models cluster is the use of non-observations Questions which ask students to rely on mathematical or physical reasoning rather than observations. While the ultimate goal of achieving a new mental model may be similar to activities in the Learning Concepts cluster, the methods often differ. In some cases, it may be that the concepts involved do not lend themselves to direct observation but are more accessible to a mathematical approach. In other cases, it may be that the underlying goal is for students to develop the sense-making, metacognitive, and epistemological skills they need to evaluate their framework of ideas. The designers emphasize conceptually complex problems which may require more of these skills. This process of developing sense-making, metacognitive, and epistemological skills and applying them to more complex situations appears to be the main motivation for the prevalence of the producing spoken representations featured in the Building Models cluster.

4.3 Conclusions

To assist instructors who want to develop or modify their activities associated with their goals by using the features of different research-based activities, we coded student actions in sixty-six activities from eleven research-based curricula and analyzed code frequencies using k-means cluster analysis. The result of this analysis was three clusters.

1) Thinking like a Scientist cluster's most important features are designing experiments by students, spoken representation, error analysis, reasonableness of student's answers, assumptions, simplifications and limitations. These activities focus attention on students performing scientific practices. These features are supported by design principles that focus mainly on designing experiments and evaluating the results.

2) Learning Concepts cluster mainly concentrates on observation and prediction. Activities in this cluster emphasize conceptual understanding of students and collecting data from experiments to test their hypothesis. This cluster uses essential points of conceptual change

Cluster Name	Cluster Description
Thinking like a Scientist	Activities in this cluster focus on students performing scientific practices such as experiment design, reaching decisions by collaboration, and examining results and processes for accuracy (such as error analysis, checking assumptions and simplifications, and evaluating reasonableness)
Learning Concepts	The most frequent features observed were Prediction and Observation. The goals expressed by the designers of activities in the Learning Concepts cluster appear consistent with the key tenets of conceptual change theory. The most frequent features show an emphasis toward students expressing their conceptual understanding and collecting data to test their ideas.
Building Models	Compared with the other clusters, the most frequent features observed in this cluster were Spoken Representation and Non-observation Questions. Non-observation Questions are physics or math questions that do not use data from a previous measurement or observation and are not prediction questions. They tend to engage students in problem-solving, refining their intuitions, using and interpreting representations, and model building, e.g. students try to prove a formula or make a hypothesis.

Table 4.6 List of the clusters' names and their properties.

theory.

3) Building Models cluster focuses on tools for helping students to solve conceptually complex problems. It's two most prominent features are spoken representation and non-observation questions where students address physics or math questions without using collected data or observations. Activities in this cluster tend to engage students in problem-

solving, refining their intuitions, using and interpreting representations, and model building, e.g. students try to prove a formula or make a hypothesis.

In this work we have identified connections between features that appear in physics activities and the goals of the designers. Making explicit the connections between the design goals and the activity features may provide instructors with a better way to select from among published activities and also lay out a clearer path to create new activities to address the learning goals they have for their students. In this way instructors may be able to create physics activities with a more consistent design philosophy.

4.4 Study limitations

One limitation of this study lies in the issue of learning cycles. Activity designers in some cases order their activities to build skills over a sequence. In this study the unit of evaluation for the cluster analysis was individual activities which were chosen randomly from the available materials from each research-based curriculum. This investigation was not designed to capture skill-building on longer scales. For example, designers of UMI stated in interviews that the learning cycle consists of a unit of instruction rather than a single activity. They state that "modeling is definitely slower, and you have to make a lot of choices of like, what content coverage versus, like, breadth versus depth."

4.5 Diverse design goals for the design groups

As mentioned earlier, all activities will do each of the features of clusters to some extent. However, some activities are in the boundary between two clusters. It means that mathematically they belong to one cluster, but their features are close to the other clusters. As you see in the Figure 4.2, three of the ISLE labs are closer to the center of Thinking like a Scientist cluster, and three others are in the Learning Concepts cluster. However, ISLE Lab 8: Reflection and Mirrors belongs to the Learning Concepts cluster but has some Thinking like a Scientist cluster features. ISLE Lab 2: One-dimensional kinematics follow a different manner. It belongs to Thinking like a Scientist cluster but has Learning Concepts features. This tendency to belong to one cluster and simultaneously close to the other cluster is not about ISLE labs. Other design groups have the same behavior, such as one of the Cornell Thinking Critically in Physics Labs (CL) Version 2018 as illustrated in Figure 4.3, three labs from SCL as illustrated in Figure 4.6, one lab from RTP as illustrated in Figure 4.5, one lab from SDI as illustrated in Figure 4.7, one lab from the TTU as illustrated in Figure 4.9, one tutorial from OST as illustrated in Figure 4.4, and finally two tutorials from TIP as illustrated in Figure 4.8. These behaviors show that when designers design these activities, they intend to cover more goals than the other activities in the same design group. Moreover, it shows that if a designer mentions specific goals in their literature does not mean that these goals should be distributed equally in each activity. According to designers' experience and circumstances, they may change the weight of goals in their activities. They are hidden facts that maybe instructors cannot find in literature or workshops. As Henderson *et al.* (30) mention, PER community should provide additional advice to faculties who want to use research-based instructional strategies (RBIS) before they begin. Providing extra support through the initial stage of using RBIS can increase the chance of successful use. Another

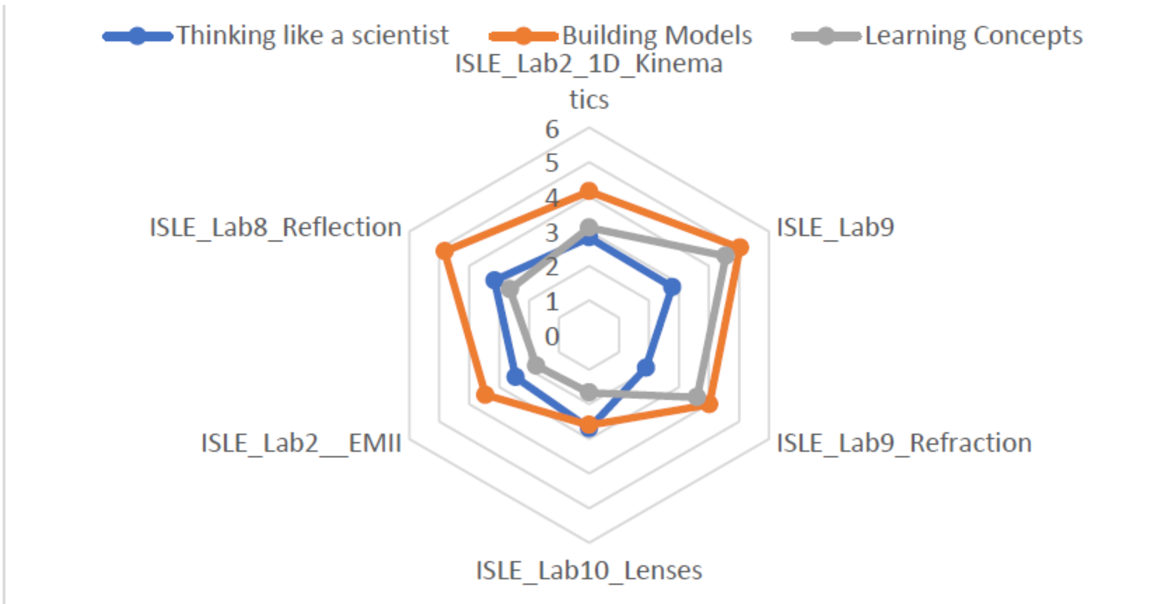

Figure 4.2 Distance from center of each cluster for ISLE labs.

point of view about the above graphs is the distance from the center of each cluster graph is to analyze groups of activities instead of concentrating on one activity. As you see in the Figures in the CL as illustrated in Figure 4.3, RTP as illustrated in Figure 4.5, UI as illustrated in Figure 4.10, SDI as illustrated in Figure 4.7, WP as illustrated in Figure 4.12, and TTU as illustrated in Figure 4.9, there is less overlap among activities than the other design groups. Maybe designers of these activities intend to follow specific goals compare to other goals.

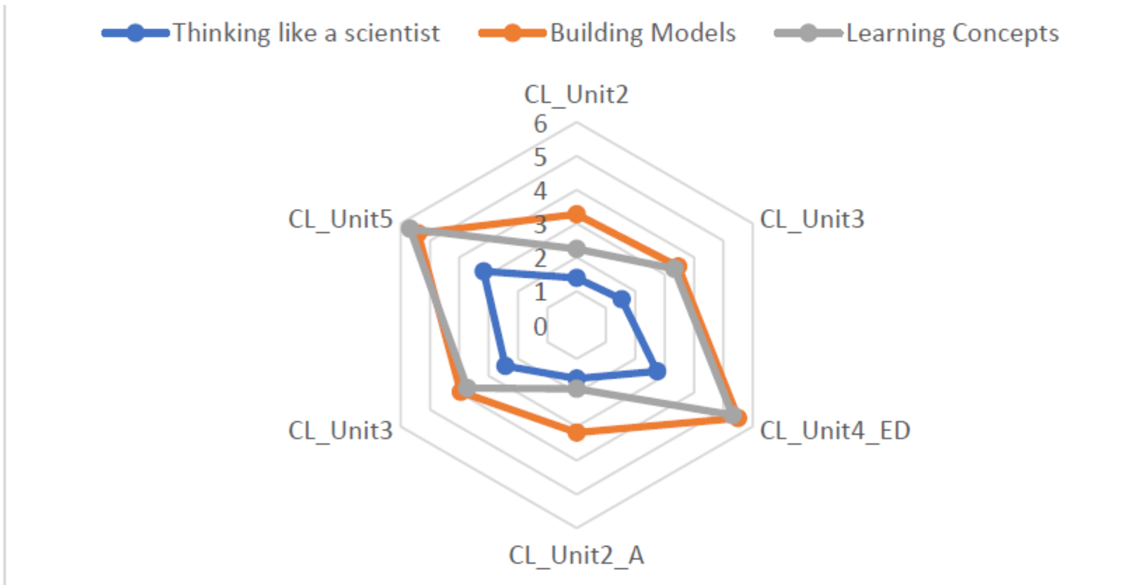

Figure 4.3 Distance from center of each cluster for Cornell Thinking Critically in Physics Labs (CL) Version 2018.

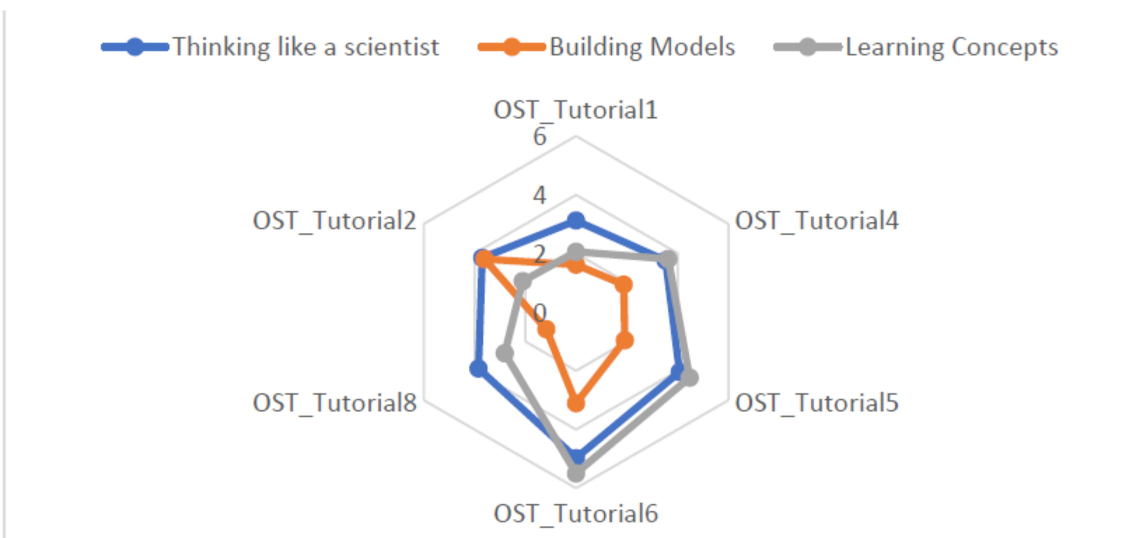

Figure 4.4 Distance from center of each cluster for Open Source Tutorials.

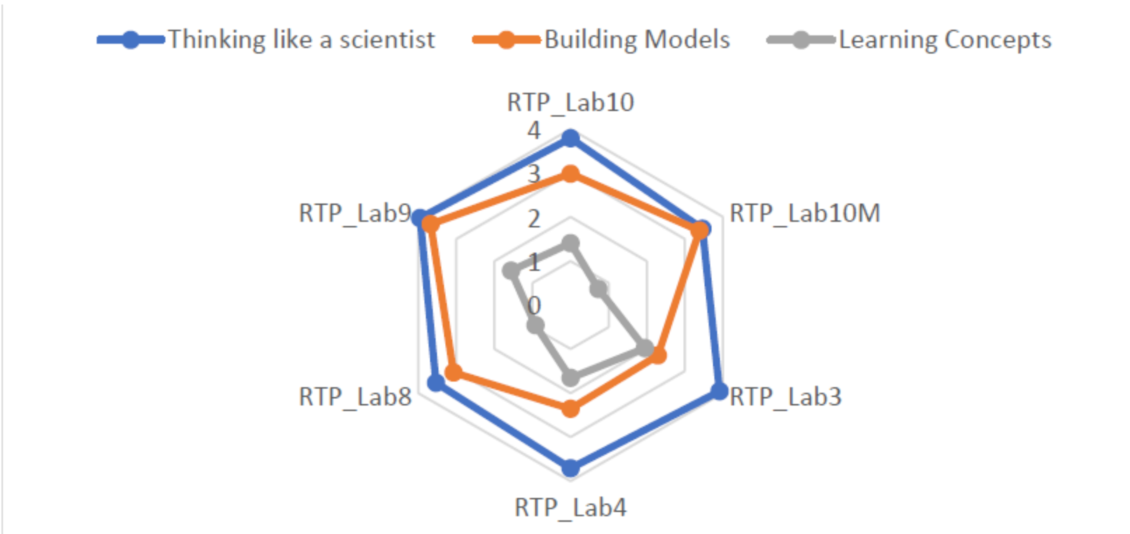

Figure 4.5 Distance from center of each cluster for RealTime Physics labs.

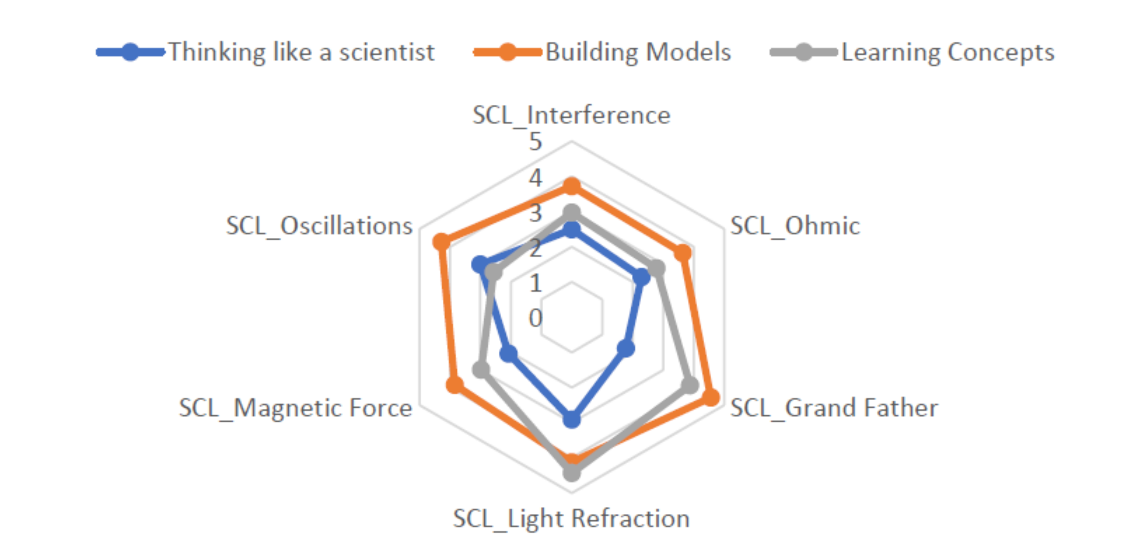

Figure 4.6 Distance from center of each cluster for Scientific Community Laboratories.

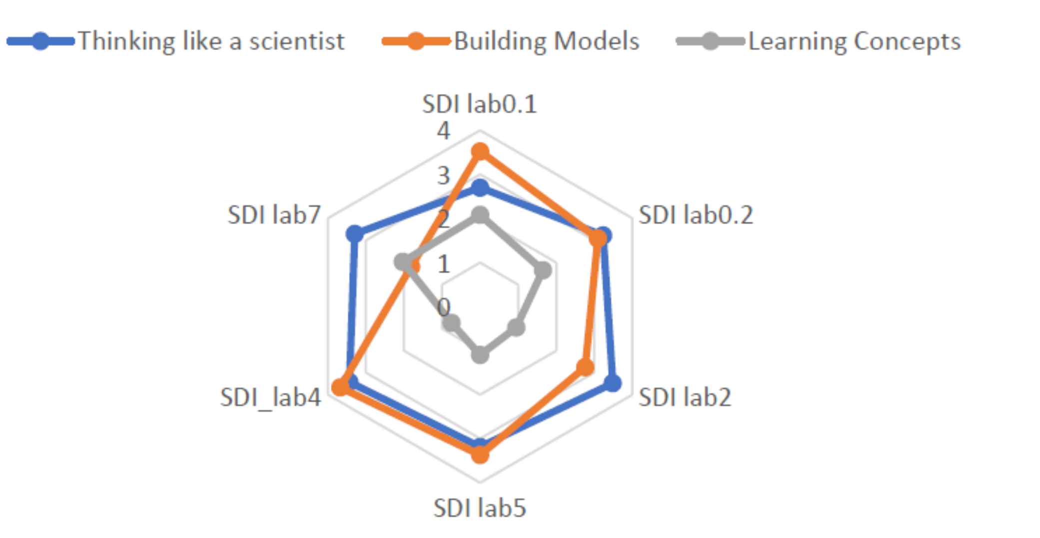

Figure 4.7 Distance from center of each cluster for Socratic Dialog Inducing Laboratories.

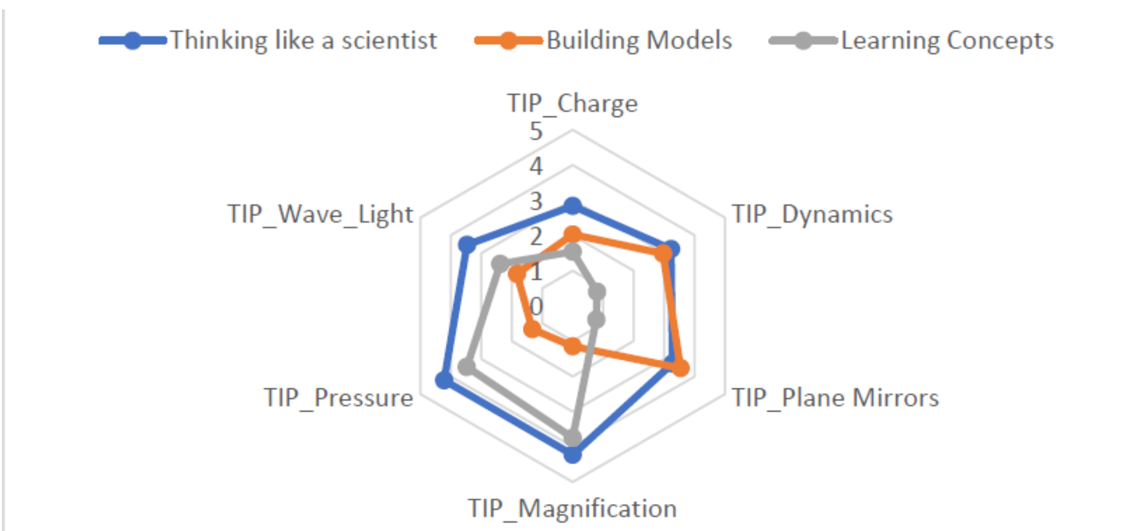

Figure 4.8 Distance from center of each cluster for Tutorials in Introductory Physics.

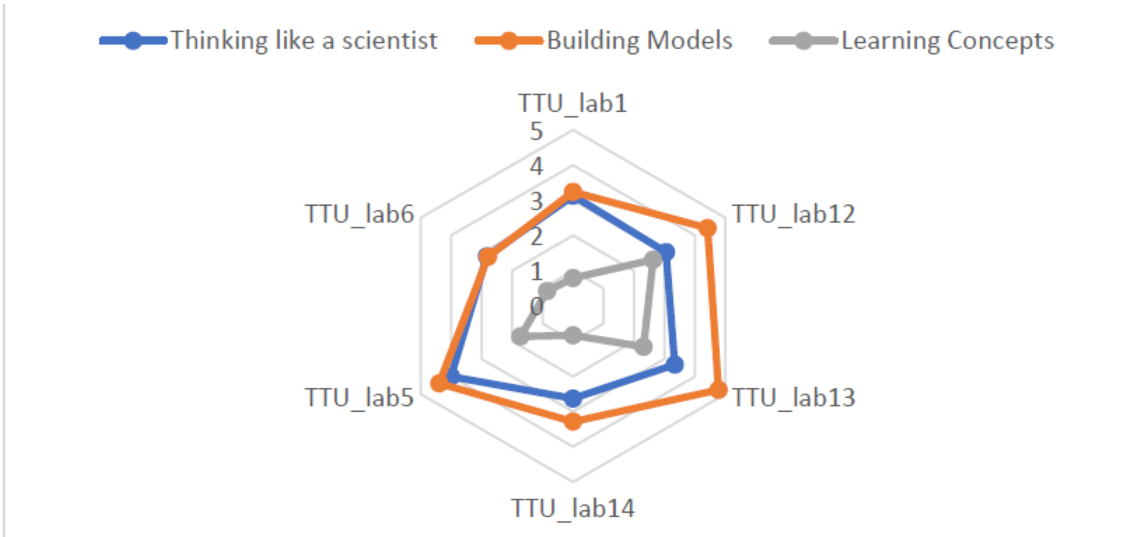

Figure 4.9 Distance from center of each cluster for Physics Department, Texas Tech University labs.

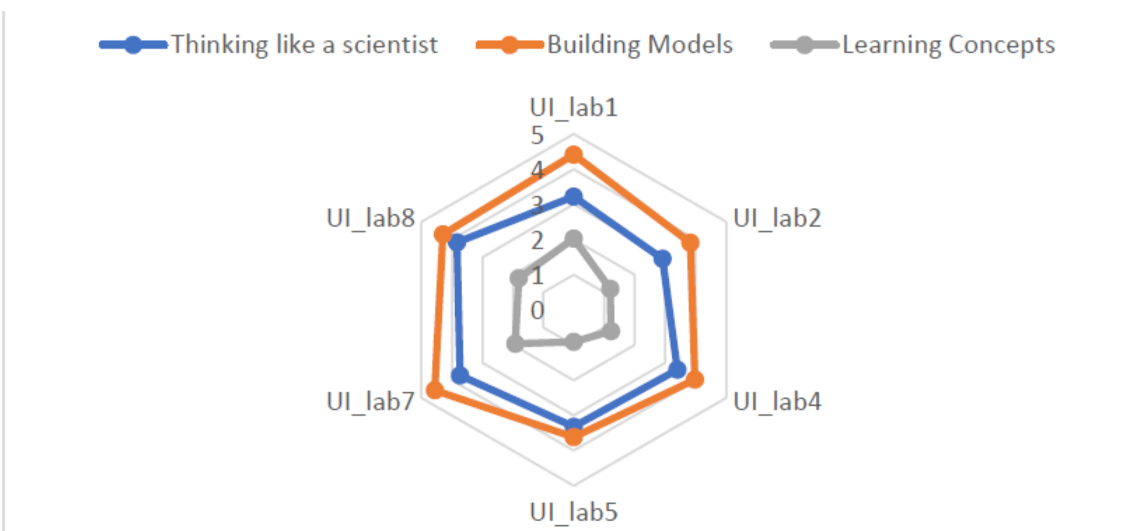

Figure 4.10 Distance from center of each cluster for University of Illinois labs.

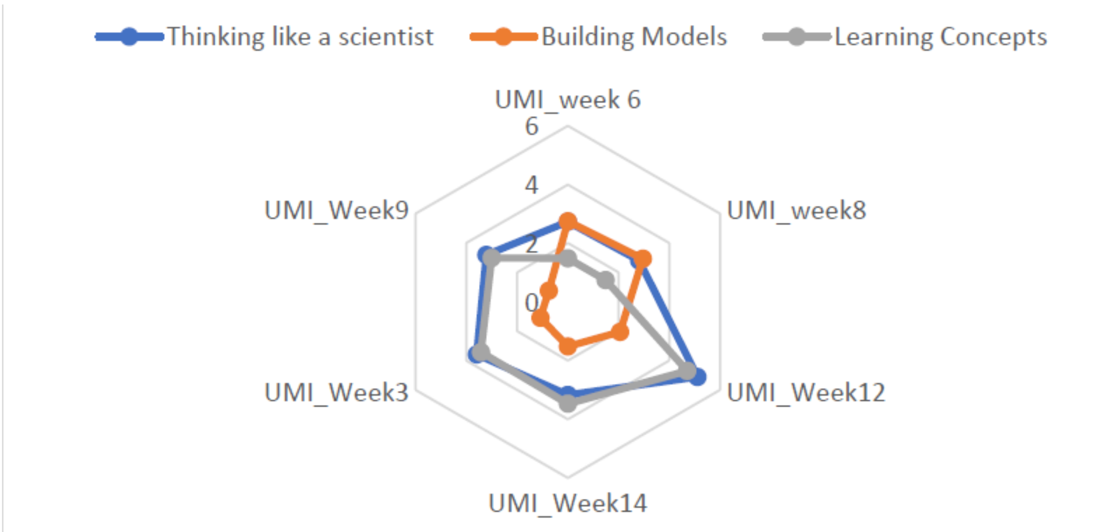

Figure 4.11 Distance from center of each cluster for University Modeling Instruction activities.

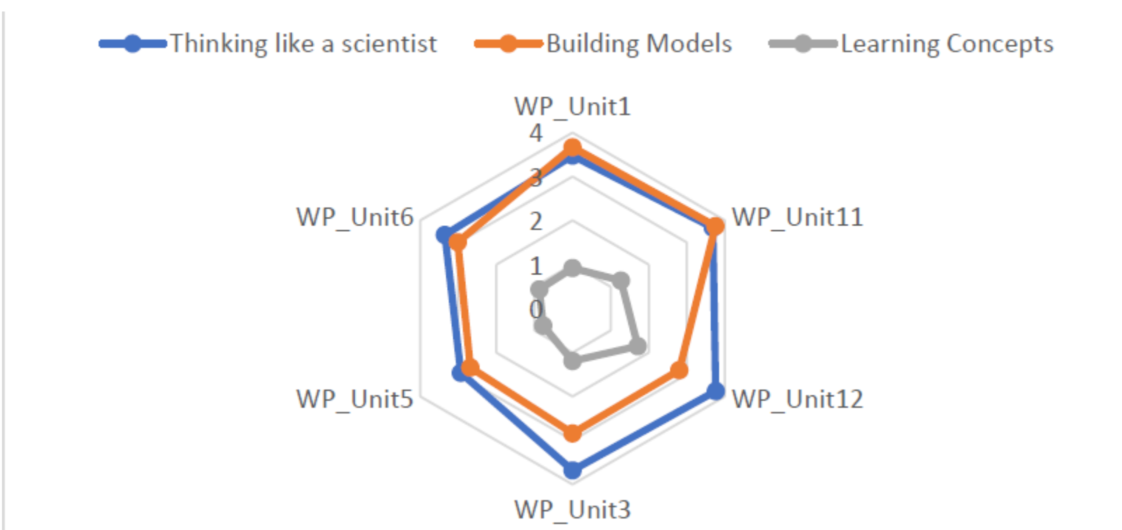

Figure 4.12 Distance from center of each cluster for Workshop Physics activities.

CHAPTER 5

Follow-up studies

5.1 Introduction

There is more information and meaning contained in the coding data than revealed by the k-means cluster analysis alone. Finding additional applications of the coding data has been the focus of follow-up studies. The following sections are a summary of my posters and talks in different national meetings. In the following sections, I analyzed more information about design and representation categories.

5.2 Designing experiments

One of the interesting features of research-based activities is design category. As I mentioned in Table 3.3 it consists of five codes.

1) Designing procedure, could include describing an experimental procedure invented by the students, explaining how the students invented an experimental process, or explaining what decisions the students had to make to invent the experimental procedure.

2) Asks students to improve their previous designs.

3) Asks students to choose an open-ended inquiry question.

4) Asks students to design/state/invent/improve a mathematical or quantitative procedure they will use before they use it.

5) Asks students to make a hypothesis that they have devised.

Design features are used more frequently in the Thinking like a Scientist cluster than in

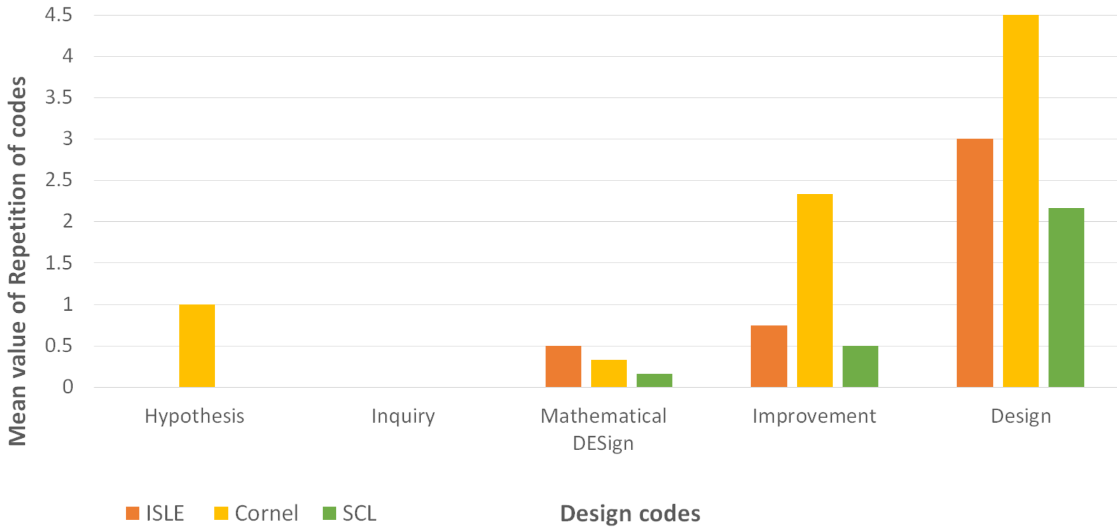

Figure 5.1 Mean value of repetition of design codes in each design group in the Thinking like a Scientist cluster.

the other clusters. For that reason we analysed mean value of repetition of individual codes in activities in this cluster as illustrated in Figure 5.1.

According to the Figure 5.1, two popular codes for student design tasks among research-based activities are design and improvement. Making hypothesis is mainly used by CL. Also, designers of physics activities are more interested in using guided inquiry compared to open inquiry.

5.3 Representations

Activities in each cluster emphasize specific representations to achieve their designers' goals. To determine which representations are associated with the design goals of each cluster we calculated the frequencies of the individual codes in the representation category for each cluster, as illustrated in Figure 5.2. The Thinking like a scientist cluster uses student-chosen

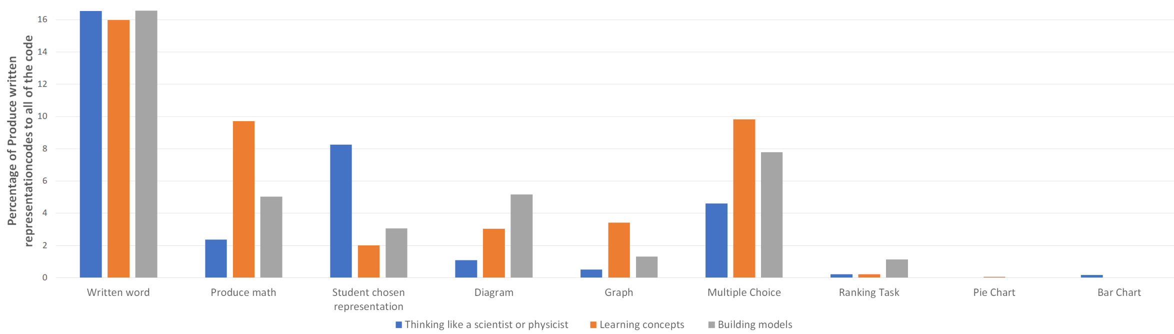

Figure 5.2 Distribution of representation codes in each cluster.

representations and written words more than other codes. Mathematical representations, multiple choices questions, and graphs were used more in Learning concepts. Building Models cluster uses more diagram representations than other clusters.

The main reason behind using student chosen representations in Thinking like a Scientist cluster is that students are asked to design experiments with less guidance than the other clusters. They are often free to choose their own representation to convey their ideas. The Learning Concepts cluster mainly uses prediction strategies. One of the questions used in this strategy is comparing experimental results with students' predictions. In this kind of question, activities ask students if their result is consistent with their prediction (yes or no questions). This is the reason multiple choices questions appear more frequently in this cluster. Also, mathematical representations and graphs are tools that designers of these activities use to convey concepts to students. In the Building Models cluster, activities guide students to make models. Students need a tool to connect their ideas to make models. Our analysis reveals that designers consider diagrams as suitable tools for this strategy.

5.4 Conclusion

We found that each cluster emphasizes specific representations to achieve their educational goals. Also, we found that activities in the Thinking like a Scientist cluster use design strategies differently than the other clusters to achieve their learning goals. Overall, we also observe that guided inquiry questions are more frequent than open inquiry questions.

CHAPTER 6
Future directions

6.1 Introduction

To apply the results of my research, we can use features of one of the clusters and redesign physics activities to determine if we can achieve the design goals representative of that cluster. Then we can use assessment tests and compare pre-test and post-test results of both redesigned and original activities to see any change in students' learning.

6.2 Designing introductory-level physics courses

If my goal is to design physics activities that can enhance students' scientific skills, I can select my activities from the "Thinking like a Scientist" cluster. To measure my success, I can use assessments such as the Physics Lab Inventory of Critical Thinking (PLIC) developed by N.G. Holmes *et al.* (75) to measure how redesigned activities can help students learn or enhance their scientific thinking skills. In particular, this assessment can measure lab skills such as comparing measurements with uncertainty, evaluating data fitted to a model, generating and evaluating conclusions based on data, designing and evaluating experimental methods.

The main components of "Building Models" are spoken representation and physics or math questions without using collected data or observations. Activities in this cluster tend to engage students in problem-solving, refining their intuitions, using and interpreting representations, model building, students trying to prove a formula, or making a hypothesis.

There is no assessment test to measure all of them, but I can use three different assessment tests according to what part of modeling they want to measure. For example, for measuring Content knowledge, they can use Mathematical Modeling Conceptual Evaluation (MMCE) developed by R. Thornton (76).

Finally, if my goal will be to design physics activities that help students learn concepts, I can choose them from physics activities in the Learning Concepts clusters. For measuring my success, then I can use different assessment tests designed for measuring Learning Concepts such as the Force Concept Inventory (FCI) (77) and Force and Motion Conceptual Evaluation (FMCE) (78).

Interviewing with instructors and students is another research plan that researchers can use to inquire how my research can be beneficial for redesigning activities. I can focus on specific student groups in the classes and register their conversations and reaction to the redesign activities. I can observe a class by taking notes or can record data with different equipment. In the next step, they can use observation protocols, such as Classroom Observation Protocol for Undergraduate STEM (COPUS) (79) or Reformed Teaching Observation Protocol (RTOP) (80). In case that no observation protocol is completely fitted with the idea of the researchers, we can analyze our observation and convert them to codes and use our codebook to suggest a new protocol. The next step is to interview with groups and individual students and then compare their answers with instructors to see how students think and feel about the activities and learn the instructor's education goals. We can involve undergraduate students in all these processes after they pass IRB courses.

REFERENCES

[1] Chandralekha Singh. What can we learn from per: Physics education research? *The Physics Teacher*, 52(9):568–569, 2014.

[2] Lillian Christie McDermott. Oersted medal lecture 2001:"physics education research—the key to student learning". *American Journal of Physics*, 69(11):1127–1137, 2001.

[3] FK Richtmyer. Physics is physics. *American Journal of Physics*, 1(1):1–5, 1933.

[4] R Beichner. An introduction to physics education. *C. Henderson, and K. e. Harper, Getting started with PER. College Park, MD: American asociation of Physics Teachers*, 2009.

[5] Del Siegle. Qualitative research. https://researchbasics.education.uconn.edu.

[6] Kathy Charmaz. Constructing grounded theory (introducing qualitative methods series). *Newcastle: SAGE.*, 2014.

[7] Eugenia Etkina, David Brookes, and Alan Van Heuvelen. Isle - investigative science learning environment. http://www.islephysics.net/index.php.

[8] Juliet Corbin and Anselm Strauss. *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage publications, 2014.

[9] Rael T. Lange. *Inter-rater Reliability*, pages 1348–1348. Springer New York, New York, NY, 2011.

[10] Julius Sim and Chris C Wright. The kappa statistic in reliability studies: use, interpre-

tation, and sample size requirements. *Physical therapy*, 85(3):257–268, 2005.

[11] Werner Vach. The dependence of cohen's kappa on the prevalence does not matter. *Journal of clinical epidemiology*, 58(7):655–661, 2005.

[12] Sarah McKagan, Adrian Madsen, and Eleanor Sayre. Physport: Supporting physics teaching with research-based resources. https://www.physport.org/.

[13] Joshua Von Korff, Amin Bayat Barooni, Hannah Pamplin, and Jacquelyn Chini. The "revisiting" strategy in physics tutorials. In *Physics Education Research Conference 2016*, PER Conference, pages 372–375, Sacramento, CA, July 20-21 2016.

[14] Joseph Kozminski, Heather Lewandowski, Nancy Beverly, Steve Lindaas, Duane Deardorff, Ann Reagan, Richard Dietz, Randy Tagg, M EblenZayas, J Williams, et al. Aapt recommendations for the undergraduate physics laboratory curriculum. *American Association of Physics Teachers*, 29, 2014.

[15] Joshua Von Korff, Amin Bayat Barooni, Monica K. Cook, Brian Ferguson, and Kyle Simmons. Design strategies for research-based physics activities. In *Physics Education Research Conference 2017*, PER Conference, pages 412–415, Cincinnati, OH, July 26-27 2017.

[16] Richard R Hake. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66(1):64–74, 1998.

[17] Robert J Beichner, Jeffery M Saul, Rhett J Allain, Duane L Deardorff, and David S Abbott. Introduction to scale-up: Student-centered activities for large enrollment uni-

versity physics. *ERIC*, 2000.

[18] Robert J Beichner, Jeffery M Saul, David S Abbott, Jeanne J Morse, Duane Deardorff, Rhett J Allain, Scott W Bonham, Melissa H Dancy, and John S Risley. The student-centered activities for large enrollment undergraduate programs (scale-up) project. *Research-based reform of university physics*, 1(1):2–39, 2007.

[19] Eric Brewe. Modeling theory applied: Modeling instruction in introductory physics. *American Journal of physics*, 76(12):1155–1160, 2008.

[20] Priscilla W Laws. Millikan lecture 1996: Promoting active learning based on physics education research in introductory physics courses. *American Journal of Physics*, 65(1):14–21, 1997.

[21] Priscilla W Laws, Maxine C Willis, and David R Sokoloff. Workshop physics and related curricula: A 25-year history of collaborative learning enhanced by computer tools for observation and analysis. *The Physics Teacher*, 53(7):401–406, 2015.

[22] David R Sokoloff, Ronald K Thornton, and Priscilla W Laws. *RealTime Physics: Active Learning Laboratories, Module 1: Mechanics*. John Wiley & Sons, 2011.

[23] David R Sokoloff, Priscilla W Laws, and Ronald K Thornton. Realtime physics: active learning labs transforming the introductory laboratory. *European Journal of Physics*, 28(3):S83, 2007.

[24] Richard Hake. Helping students to think like scientists in socratic dialogue-inducing labs. *The Physics Teacher*, 50(1):48–52, 2012.

[25] Richard Hake. Socratic dialog-inducing (sdi) labs.

https://www1.physics.indiana.edu/ sdi/.

[26] Richard R Hake. Socratic pedagogy in the introductory physics laboratory. *The physics teacher*, 30(9):546–552, 1992.

[27] NG Holmes and DA Bonn. Quantitative comparisons to promote inquiry in the introductory physics lab. *The Physics Teacher*, 53(6):352–355, 2015.

[28] NG Holmes, Carl E Wieman, and DA Bonn. Teaching critical thinking. *Proceedings of the National Academy of Sciences*, 112(36):11199–11204, 2015.

[29] Rebecca Lippmann. *Students' Understanding of Measurement and Uncertainty in the Physics Laboratory: Social construction, underlying concepts, and quantitative analysis.* Doctoral dissertation, University of Maryland, January 2003.

[30] Charles Henderson, Melissa Dancy, and Magdalena Niewiadomska-Bugaj. Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Physical Review Special Topics-Physics Education Research*, 8(2):020104, 2012.

[31] Charles Henderson and Melissa H Dancy. Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations. *American Journal of Physics*, 76(1):79–91, 2008.

[32] David E Meltzer and Ronald K Thornton. Resource letter alip–1: active-learning instruction in physics. *American journal of physics*, 80(6):478–496, 2012.

[33] Eugenia Etkina. Millikan award lecture: Students of physics—listeners, observers, or collaborative participants in physics scientific practices?, 2015.

[34] Beth Thacker, Hani Dulli, Dave Pattillo, and Keith West. Lessons from a large-scale assessment: Results from conceptual inventories. *Physical Review Special Topics-Physics Education Research*, 10(2):020104, 2014.

[35] Natasha G Holmes and Carl E Wieman. Introductory physics labs: We can do. *Physics Today*, 71:1–38, 2018.

[36] Eugenia Etkina and Gorazd Planinšič. Thinking like a scientist. *Physics world*, 27(03):48, 2014.

[37] Benjamin M Zwickl, Noah Finkelstein, and HJ Lewandowski. Incorporating learning goals about modeling into an upper-division physics laboratory experiment. *American Journal of Physics*, 82(9):876–882, 2014.

[38] Bethany R Wilcox and Heather J Lewandowski. Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students' beliefs about experimental physics. *Physical Review Physics Education Research*, 13(1):010108, 2017.

[39] Joshua Von Korff, Chang Zhan, Birjoo Vaishnav, Jacquelyn Chini, Ashley Warneke, and Ozden Sengul. The use of representations in evidence-based and non-evidence-based physics activities. In *Physics Education Research Conference 2015*, PER Conference, pages 351–354, College Park, MD, July 29-30 2015.

[40] Amin Bayat Barooni, Joshua Von Korff, Brian D Thoms, Zeynep Topdemir, and Jacquelyn J Chini. Three research-based approaches to the design of physics activities. *arXiv preprint arXiv:2009.04020*, 2020.

[41] Michael Quinn Patton. Enhancing the quality and credibility of qualitative analysis.

Health services research, 34(5 Pt 2):1189, 1999.

[42] Hervé Abdi and Lynne J Williams. Encyclopedia of research design. *Coefficient of Variation*, 2010.

[43] Joshua Von Korff, Benjamin Archibeque, K Alison Gomez, Tyrel Heckendorf, Sarah B McKagan, Eleanor C Sayre, Edward W Schenk, Chase Shepherd, and Lane Sorell. Secondary analysis of teaching methods in introductory physics: A 50 k-student study. *American Journal of physics*, 84(12):969–974, 2016.

[44] David Sokoloff. Curricula for active learning in the introductory physics course. https://pages.uoregon.edu/sokoloff/Active$_{L}$*earning.html*.

[45] Eugenia Etkina, Gorazd Planinšič, and Michael Vollmer. A simple optics experiment to engage students in scientific inquiry. *American journal of physics*, 81(11):815–822, 2013.

[46] Priscilla Laws, Robert Boyle, Patrick Cooney, Kenneth Laws, John Luetzelschwab, David Sokoloff, and Ronald Thornton. Workshop physics. https://www.physport.org/methods/method.cfm?G=Workshop$_{P}$*hysics*.

[47] Eric Brewe. University modeling instruction. http://univ-modelinginstruction.com/.

[48] Joe Redish. Tutorials from the university of maryland physics education resaerch group. http://umdperg.pbworks.com/w/page/10511238/Tutorials.

[49] Kevin A Hallgren. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1):23, 2012.

[50] Brian S Everitt. *Making sense of statistics in psychology: A second-level course.* Oxford

University Press, 1996.

[51] Joseph L Fleiss, Bruce Levin, Myunghee Cho Paik, et al. The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2(212-236):22–23, 1981.

[52] Columbia University's Mailman School of Public Health. K-means cluster analysis. https://www.publichealth.columbia.edu/research/population-health-methods/k-means-cluster-analysis.

[53] Analytics Vidhya. Getting your clustering right (part i). https://www.analyticsvidhya.com/blog/2013/11/getting-clustering-right/.

[54] Junjie Wu. Cluster analysis and k-means clustering: an introduction. In *Advances in K-means Clustering*, pages 1–16. Springer, 2012.

[55] Anton K Formann. *Die latent-class-analyse: Einführung in Theorie und Anwendung.* Beltz, 1984.

[56] John A Hartigan and Manchek A Wong. Ak-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(1):100–108, 1979.

[57] Robert L Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, 1953.

[58] Rachel E. Scherr and Andrew Elby. Open source tutorials. https://www.physport.org/methods/method.cfm?G=OST$_{T}utorials$.

[59] Eric Brewe. University modeling instruction. http://univ-modelinginstruction.com/.

[60] Priscilla Laws, Boyle Robert, Patrick Cooney, Kenneth Laws, John Luetzelschwab, David Sokoloff, and Ronald Thornton. Workshop physics. here.

[61] Lillian C. McDermott, Peter S. Shaffer, and the Physics Ed-

ucation Group at UW. Tutorials in introductory physics. https://www.physport.org/methods/method.cfm?G=Tutorials$_{i}n_{I}ntroductory_{P}hysics$.

[62] David Sokoloff, Ron Thornton, and Priscilla Laws. Realtime physics. https://www.physport.org/methods/method.cfm?G=Realtime$_{P}hysics$.

[63] Eugenia Etkina and Alan Van Heuvelen. Investigative science learning environment. https://www.physport.org/methods/method.cfm?G=ISLE.

[64] Rebecca Lippmann Kung, Paul Gresser, and Joe Redish. Scientific community labs. https://www.physport.org/methods/method.cfm?G=SCL.

[65] Priscilla Laws. Workshop physics: Learning introductory physics by doing it. *Change: The Magazine of Higher Learning*, 23(4):20–27, 1991.

[66] Mila Kryjevskaia, Andrew Boudreaux, and Dustin Heins. Assessing the flexibility of research-based instructional strategies: Implementing tutorials in introductory physics in the lecture environment. *American Journal of Physics*, 82(3):238–250, 2014.

[67] Eugenia Etkina and Alan Van Heuvelen. Investigative science learning environment–a science process approach to learning physics. *Research-based reform of university physics*, 1(1):1–48, 2007.

[68] Eugenia Etkina, Alan Van Heuvelen, David T Brookes, and David Mills. Role of experiments in physics instruction—a process approach. *The Physics Teacher*, 40(6):351–355, 2002.

[69] Wilson J González-Espada, Jennifer Birriel, and Ignacio Birriel. Discrepant events: A challenge to students' intuition. *The Physics Teacher*, 48(8):508–511, 2010.

[70] Michelene TH Chi. Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. *International handbook of research on conceptual change*, 61:82, 2008.

[71] Claudia Khourey-Bowers. Active learning strategies: The top 10. *The Science Teacher*, 78(4):38, 2011.

[72] Joseph Hesse. From naive to knowledgeable. *The Science Teacher*, 56(6):55–58, 1989.

[73] David Hestenes. Toward a modeling theory of physics instruction. *American journal of physics*, 55(5):440–454, 1987.

[74] Laura Lising and Andrew Elby. The impact of epistemology on learning: A case study from introductory physics. *American Journal of Physics*, 73(4):372–382, 2005.

[75] Cole Walsh, Katherine N Quinn, C Wieman, and Natasha G Holmes. Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking. *Physical Review Physics Education Research*, 15(1):010135, 2019.

[76] Ronald K Thornton. Measuring and improving student mathematical skills for modeling. In *Proceedings GIREP Conference*, 2006.

[77] David Hestenes, Malcolm Wells, and Gregg Swackhamer. Force concept inventory. *The physics teacher*, 30(3):141–158, 1992.

[78] Ronald K Thornton and David R Sokoloff. Assessing student learning of newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *american Journal of Physics*, 66(4):338–352, 1998.

[79] Michelle K Smith, Francis HM Jones, Sarah L Gilbert, and Carl E Wieman. The

classroom observation protocol for undergraduate stem (copus): A new instrument to characterize university stem classroom practices. *CBE—Life Sciences Education*, 12(4):618–627, 2013.

[80] Daiyo Sawada, Michael D Piburn, Eugene Judson, Jeff Turley, Kathleen Falconer, Russell Benford, and Irene Bloom. Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School science and mathematics*, 102(6):245–253, 2002.