Examining Sources of Variation in Student Confusion in College Classes

Youjie Chen
Stanford University
Stanford, CA
minachen@stanford.edu

René F. Kizilcec
Cornell University
Ithaca, NY
kizilcec@cornell.edu

ABSTRACT
Students often experience confusion while learning, and if promptly resolved, it can promote engagement and deeper understanding. However, detecting student confusion and intervening in a timely and scalable manner challenges even seasoned instructors. To understand when and where students are most likely to be confused, we study the systematic occurrence of confusion in college classes among 29,511 students in twelve universities. We use a novel method for affect detection that allows students to self-report confusion on individual presentation slides during their classes. Across 1,366 class presentations, we find that confusion arises at different times during class and depends on class duration, class size, type of institution, and academic discipline. Confusion is most prevalent during short presentations, in small classes, low-tier institutions, and scientific disciplines.

Author Keywords
Confusion; Affect Detection; PowerPoint; Instructional Design; Higher Education

CCS Concepts
•Human-centered computing → Empirical studies in HCI;
•Applied computing → Education;

INTRODUCTION
Confusion is one of the most frequent affective states in the learning process [7]. If appropriately resolved, the experience of confusion can be beneficial to learners in fostering engagement and developing a deeper level of comprehension [10]. However, if learners fail to resolve confusion after a period of time, it can give way to frustration and boredom, and be detrimental to learning outcomes [8]. This presents a common problem for classes where the instructor cannot recognize and respond to student confusion in a timely manner. Advancing an understanding of when confusion is likely to arise in real-world class settings can help instructors and instructional designers better support student learning.

L@S '20, August 12–14, 2020, Virtual Event, USA.

ACM ISBN 978-1-4503-7951-9/20/05 ...$15.00.
http://dx.doi.org/10.1145/3386527.3405939

A large number of studies have investigated ways to detect confusion in digital learning environments, such as through mouse-clicks [11], eye-tracking [16], or language in discussion posts [19]. However, only a few have explored ways to detect confusion in mainstream in-person classes [13, 15]. Classroom instructors often rely on signals such as facial expressions and raising hands to recognize student confusion. However, these traditional strategies do not scale well to large class sizes and cultures in which students prefer not to express confusion in front of others.

Educators have developed techniques to obtain more reliable measures of student confusion. One technique is to equip students with a clicker so that instructors can pause and poll student confusion during class [5, 4]. However, adoption of clickers as an additional device in large classrooms can still be troublesome and polling students in the middle of class is a form of task switching that interrupts the flow of learning. A more advanced technique is to install physical sensors in classrooms, such as facial recognition cameras or multi-modal sensors to detect confusion [9]. However, such technology is not widely adopted, costly when applied at scale, and raises concerns about student privacy and algorithmic bias.

In this paper, we use a new technique that has students self-report confusion in a mobile application called Rain Classroom [18]. The application integrates with Microsoft PowerPoint, one of the most popular tools for in-class teaching in higher education. It allows students to anonymously express confusion about specific slides in class by pressing a button on their smartphone. This technique collects just-in-time reports of student confusion for any number of students with minimal interruption and implementation cost. We investigate the timing of student confusion in college classes across twelve Chinese universities over five semesters. We leverage detailed log data of over 1,000 presentations to explore instructors' usage of PowerPoint presentations and the heterogeneity in student confusion among different types of institutions, disciplines, and class sizes.

METHODS
We use Rain Classroom for data collection. It is a plug-in for PowerPoint that allows instructors to create presentations with several added interactive functionalities in class, such as quizzing and polling [18]. Rain Classroom has been adopted in thousands of university classrooms across China, encompassing millions of PowerPoint presentations. Instructors use their computers to create a PowerPoint presentation with the

Rain Classroom plug-in and use their smartphones to control the presentation. When an instructor starts a presentation in class, a QR code is displayed on-screen. By scanning the QR code, students connect to the presentation on their smartphone. As each individual slide is presented in class, students simultaneously see the slide on their smartphone to follow along and take notes. Students can look back at prior slides by swiping through presentations on their smartphones at any time, but they cannot view slides that have not been presented by the instructor yet.

Students can engage with individual slides on their smartphone while the instructor teaches. When a slide is shown on a student's phone, it is accompanied by two buttons: an "I don't understand" button and a "Bookmark" button. When a student clicks on the "I don't understand" button, an anonymous notification is immediately sent to the instructor's smartphone, which displays the number of students who report confusion about the slide. The instructor can either choose to adjust teaching on the fly or take notice of the notification after class. When a student clicks on the "Bookmark" button, no notification is sent to the instructor, but the slide gets bookmarked on the student's smartphone so that they can review it later. Students can undo their clicks on either button once their confusion is resolved. In this study, we only analyze instances where the click was not undone. As we note above, confusion is beneficial to learning if it can be resolved shortly afterwards, but unresolved confusion can be detrimental to learning outcomes and should therefore be avoided. This feature can be particularly useful in large classrooms where not every student has the chance to express their level of understanding in class, or in cultures where students are cautious about expressing confusion in class. We rely on this feature to operationalize our study of in-class student confusion.

Sample Selection

We focus on data collected between September 2017 and January 2020. We randomly sampled data from twelve universities that have used Rain Classroom for over a year. The number of PowerPoint presentations developed by these universities ranges from 328 to 731 (the range across the entire dataset is between dozens and thousands per institution). The twelve universities comprise four first-tier institutions, four second-tier institutions, and four third-tier institutions. The tiers are officially defined by college entrance exam scores and enrollment size according to the Chinese college entrance system. As the use of Rain Classroom is not constrained to classroom settings (it is also used in public lectures and small-scale demonstrations), we apply several filters to focus on PowerPoint presentations that are most likely to have occurred in classroom settings: (1) the presentation was presented: it has a start and end time; (2) the presentation duration is between 20–150 minutes; (3) the presentation length is between 5–205 slides; (4) the audience size is between 10–500 students.

A total of 5,957 PowerPoint presentations with 356,371 slides met all four criteria. Manual checks with random slide samples confirmed that the presentations were used for instructional purposes. We also manually checked the use of animations in presentations and found that among the presentations with animations, almost all presentations used animation within slides rather than spaced out across slides. The use of animation is therefore unlikely to confound analyses involving slide counts.

Through conversations with Rain Classroom staff and instructors, we learned that the "I don't understand" and "Bookmark" features are not used in many classrooms. Thus, the absence of clicks may be a sign of unawareness of the feature rather than understanding of the learning materials. To focus on those presentations where students have an awareness of the feature, we selected classes with at least one click of "I don't understand" and one click of "Bookmark" during a class presentation. This reduced our final sample to 1,366 PowerPoint presentations with 86,332 slides and a combined audience size of 29,511 students.

Measurement of Confusion

Students can indicate confusion during class by using their smartphone. Self-reported confusion can provide a reliable indicator of actual confusion. Multiple prior studies that have used similar approaches and used self-report as the "ground truth" for building affect detectors [4, 12, 6]. However, we also acknowledge its meta-cognitive limitations: students need to be self-aware of confusion and willing to admit it. We interviewed Rain Classroom staff and instructors to understand how the two buttons are used in practice. We learned that although clicking "I don't understand" sends a clear and anonymous message that students are confused and seeking help, not all students are comfortable sending a message to the instructor. Instead, many students "Bookmark" confusing slides so that they can review them later. Through interviews we learned that while clicks on "I don't understand" are clearly indicative of confusing slides, it overlooks many instances of student confusion. While some "Bookmark" clicks are intended for remembering key slides, some are indicative of confusing slides. The data reflects this insight, in that there are ten times more clicks on "Bookmark" than "I don't understand".

We therefore use the following definition of student confusion: a slide is deemed confusing if at least 2% of students in the class click either button, and at least 1% click "I don't understand". The rationale is that bookmarked slides are more likely to be confusing if the bookmarking co-occur with clicks on "I don't understand". When a student clicks either button, the slide index (first slide, second slide, etc.) and the time of the click are automatically recorded. We use the slide index to denote the location of student confusion. We use the median of all click times for the same slide index to denote the time of student confusion.

RESULTS

Presentation Length and Duration

Before examining student confusion, we first look at how instructors use in-class PowerPoint presentations in terms of their length (number of slides) and duration (time spent). Figure 1 visualizes the distribution of presentations along both dimensions, where lighter colors indicate more presentations of a given length in 10-slide bins and duration in 10-minute bins. We find that there are two distinct clusters of presentations in terms of their duration. One contains short presentations under

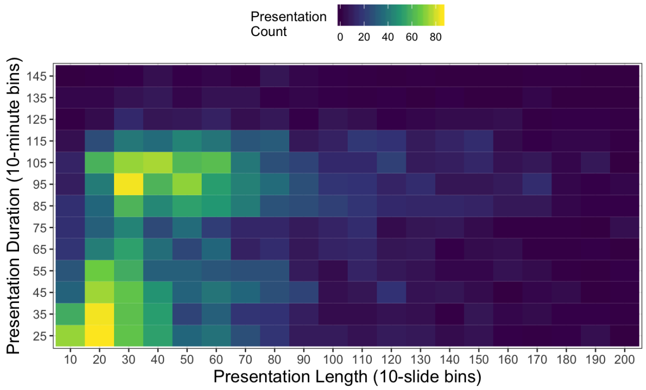

Figure 1. Distribution of PowerPoint presentations in terms of length (number of slides) and duration (time spent).

70 minutes, while the other contains long presentations above 70 minutes. Presentation duration stands out as a key distinguishing factor, probably because institutions have restrictions on class scheduling that influence the available time for presentations. As the total amount of class time may affect student confusion, we will examine how student confusion differs between the two clusters of long and short presentations.

Variation in Student Confusion

To understand when confusion is most likely to occur, we investigate several potential sources of variation: presentation duration, class size, institution type, and academic disciplines. We divided presentations into long presentations (over 70 minutes) and short presentations (under 70 minutes) based on the observed clustering. The class size in our data ranges from 10 to 500 students. We divided them into small classes (less than 50 students), medium classes (50 to 99 students), and large classes (100 to 500 students). Our presentation data comes from three types of institution, namely first-tier, second-tier, and third-tier institution. The classification of academic disciplines varies across universities and the presentations in our sample span a wide range of disciplines. We categorized all disciplines into two sets, "hard" and "soft" disciplines, based on the binary classification proposed by Biglan [2, 1, 14, 17]. Disciplines with few paradigms like mathematics and physics are classified as "hard," while disciplines with many paradigms like in the humanities are classified as "soft."

We fit a logistic mixed-effects model to predict the occurrence of confusion on a slide using the following set of categorical predictor variables as defined above: presentation duration, class size, institution type, and academic discipline. We model each presentation as random intercept. The regression output is summarized in Table 1. The coefficient estimates represent changes in the log likelihood of slide confusion from the baseline (long presentation, medium class, hard academic discipline, third-tier institution). We computed the variance inflation factor (VIF) score to test for factor collinearity and found minimal evidence for collinearity (VIF scores < 2).

The regression results indicate that students have 23% higher odds of being confused during shorter presentations than longer presentations ($z = 2.34, p = 0.02$). This pattern could

Table 1. Mixed-Effect Logistic Regression Predicting Student Confusion

Predictor Variables	Coefficient (Std. Error)
Presentation Duration = Short	0.21^{*}
	(0.09)
Class Size = Large	-0.10^{***}
	(0.11)
Class Size = Small	0.29^{*}
	(0.11)
Institution Type = First-Tier	-0.47^{***}
	(0.11)
Institution Type = Second-Tier	-0.49^{***}
	(0.10)
Academic Discipline = Soft	-0.45^{***}
	(0.09)
Constant	-3.56^{***}
	(0.10)
N	86332

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

suggest that instructors attempt to cover too much material during shorter presentations. Compared to medium-sized classes, students have 34% higher odds of being confused in small classes ($z = 2.58, p = 0.01$) and 10% lower odds of being confused in large classes ($z = -9.25, p < 0.001$). Depending on whether it is appropriately resolved or not, confusion can be an intermediary state that either gives way to persistent engagement or to disengagement. Past research shows that inattention is more prevalent in larger classrooms [3]. We suspect that the significantly lower odds of confusion in large classes is a sign of disengagement rather than clear understanding.

Comparing between institution types, there is no significant difference in confusion between first-tier and second-tier institutions ($z = -0.15, p = 0.88$), but students in third-tier institutions have 60% higher odds of confusion compared to students in first-tier institutions ($z = 4.35, p < 0.001$). This might indicate that students from third-tier institutions have less developed learning strategies on average, or instructors at these institutions have weaker teaching skills on average, which both can lead to increased confusion. For students taking courses in "soft" disciplines, the odds of confusion are 36% lower than in "hard" disciplines ($z = -4.815, p < 0.001$). According to Biglan [1], "hard" disciplines feature more instances of formal reasoning and use of symbolic languages, which may lead to more confusion.

DISCUSSION

This study presents a novel approach to detect and analyze confusion at large scale. We use an interactive in-class tool that enables students to self-report confusion during class without interruption. This approach allows us to not only capture real-time student reactions to learning materials in classrooms, but also to compare across a large number of presentation characteristics in various contexts.

Furthermore, this study contributes new insights into student confusion in relation to instructor usage of PowerPoint presentations and how confusion varies by presentation duration, class size, institution type, and academic disciplines. One surprising finding is that students report significantly more confusion in short and small classes. A possible explanation is that in these contexts, frequent confusion may be a result of increased cognitive engagement (i.e. more learning happening). This finding warrants further investigation. Another finding that stands out is that students in third-tier institutions, and "hard" disciplines are significantly more likely to encounter confusion. This can be particularly alarming as confusion can be cumulative and eventually result in frustration and disengagement [8].

Future research could investigate why students experience significantly more confusion in these contexts and how this issue might be addressed. Moreover, future studies could also examine how specific instructional interventions can be designed to mitigate sustained confusion. One possible direction is improvement of slide design. For example, disciplines related to mathematics rely heavily on inferences and reasoning, which can cause cognitive overload unless the learning materials are appropriately scaffolded. Spreading equations and formulas across several slides can alleviate cognitive load for novice learners. Other instructional interventions using new teaching technologies could help provide just-in-time feedback in class when confusion arises.

REFERENCES

[1] Anthony Biglan. 1973a. The characteristics of subject matter in different academic areas. *Journal of applied Psychology* 57, 3 (1973), 195.

[2] Anthony Biglan. 1973b. Relationships between subject matter characteristics and the structure and output of university departments. *Journal of applied psychology* 57, 3 (1973), 204.

[3] Lofter Bjarnason. 1925. Relation of class size to control of attention. *The Elementary School Journal* 26, 1 (1925), 36–41.

[4] Diane M Bunce, Elizabeth A Flens, and Kelly Y Neiles. 2010. How long can students pay attention in class? A study of student attention decline using clickers. *Journal of Chemical Education* 87, 12 (2010), 1438–1443.

[5] Jane E Caldwell. 2007. Clickers in the large classroom: Current research and best-practice tips. *CBE—Life Sciences Education* 6, 1 (2007), 9–20.

[6] M Ali Akber Dewan, Mahbub Murshed, and Fuhua Lin. 2019. Engagement detection in online learning: a review. *Smart Learning Environments* 6, 1 (2019), 1.

[7] Sidney D'Mello. 2013. A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology* 105, 4 (2013), 1082.

[8] Sidney D'Mello and Art Graesser. 2012. Dynamics of affective states during complex learning. *Learning and Instruction* 22, 2 (2012), 145–157.

[9] Toby Dragon, Ivon Arroyo, Beverly P Woolf, Winslow Burleson, Rana El Kaliouby, and Hoda Eydgahi. 2008. Viewing student affect and learning through classroom observation and physical sensors. In *International Conference on Intelligent Tutoring Systems*. Springer, 29–39.

[10] Sidney D'Mello, Blair Lehman, Reinhard Pekrun, and Art Graesser. 2014. Confusion can be beneficial for learning. *Learning and Instruction* 29 (2014), 153–170.

[11] Elena L Glassman, Juho Kim, Andrés Monroy-Hernández, and Meredith Ringel Morris. 2015. Mudslide: A spatially anchored census of student confusion for online lecture videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1555–1564.

[12] Ashish Kapoor, Winslow Burleson, and Rosalind W Picard. 2007. Automatic prediction of frustration. *International journal of human-computer studies* 65, 8 (2007), 724–736.

[13] Daniel B King. 2011. Using clickers to identify the muddiest points in large chemistry classes. *Journal of chemical education* 88, 11 (2011), 1485–1488.

[14] Gary D Malaney. 1986. Differentiation in graduate education. *Research in Higher Education* 25, 1 (1986), 82–96.

[15] Frederick Mosteller. 1989. The 'Muddiest Point in the Lecture'as a feedback device. *On Teaching and Learning: The Journal of the Harvard-Danforth Center* 3 (1989), 10–21.

[16] Mariya Pachman, Amaël Arguel, Lori Lockyer, Gregor Kennedy, and Jason Lodge. 2016. Eye tracking and early detection of confusion in digital learning environments: Proof of concept. *Australasian Journal of Educational Technology* 32, 6 (2016).

[17] Judith L Stoecker. 1993. The Biglan classification revisited. *Research in Higher Education* 34, 4 (1993), 451–464.

[18] Shuaiguo Wang and Youjie Chen. 2018. Rain classroom: a tool for blended learning with MOOCs. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. 1–2.

[19] Ziheng Zeng, Snigdha Chaturvedi, and Suma Bhat. 2017. Learner Affect through the Looking Glass: Characterization and Detection of Confusion in Online Courses. *International Educational Data Mining Society* (2017).