

HILT

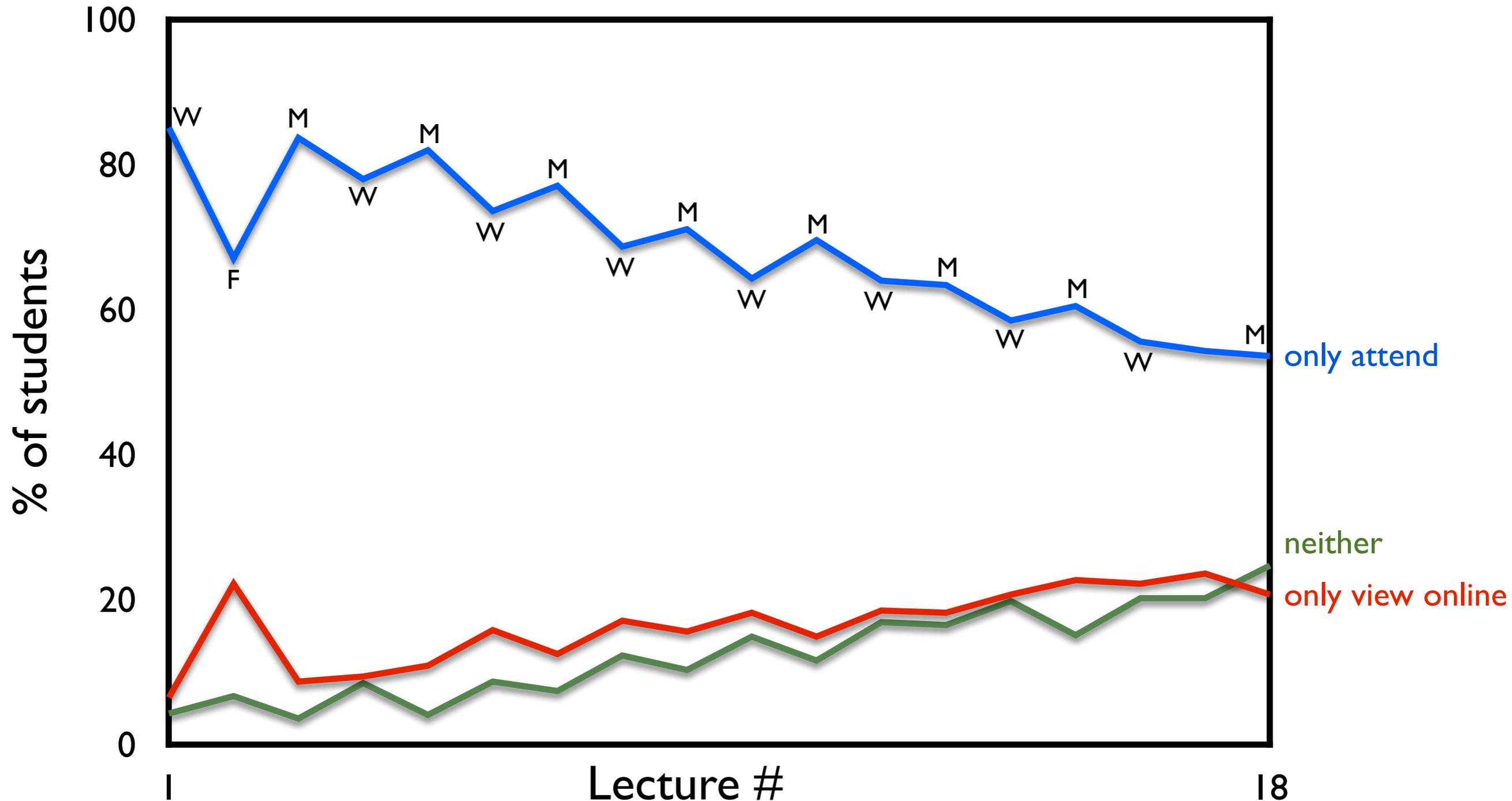
Harvard Initiative for
Learning & Teaching

Lecture attendance research: Methods and preliminary findings

November 2014



Inspiration and early findings



Inspiration for this project came from self-reported attendance and online viewing data from the 2011 iteration of CS50 (n=452). Something important changes over time that affects attendance on all time scales: from freshman to senior year (not shown), from the start to the end of the semester, and from Monday to Wednesday. By the end of this very popular course, 25% of students reported neither attending or watching lectures online. Lecture attendance was negatively correlated with online lecture viewing.



Weak extant literature



Undergraduate lecture attendance has not been comprehensively or rigorously measured. Past research is limited by one or more of the following:

- self reported attendance
- self reported reasons for attending or missing lectures
- biased sample of included courses
- single course analyses

Undergraduate lecture attendance has not been systematically related to course variables, course ratings, or the availability or use of digital resources (e.g., lecture videos).



Methods



- Four GoPro cameras installed visibly at front of four lecture halls with existing video recording equipment
- Imagery data collected every minute (slowest time lapse possible with GoPros) from all courses that met in these lecture halls
- Encrypted data retrieved regularly and stored securely
- Machine learning algorithm developed iteratively based on emerging methods from computer vision
- Imagery data paired with specific course lectures based on manual reconciliation of Harvard-accessible (behind PIN) data from course syllabi, course sites, and published meeting times/locations
- One summary statistic derived from multiple estimates per lecture
- Researcher given administrative access to course sites, which included aggregate site and lecture video usage statistics



Development of algorithm



- The algorithm was developed and refined over seven months
- No “out of box” solution existed
- Small sample of images used initially to explore different strategies for estimating attendance
- Algorithm was “trained” to distinguish between empty seats and seats with people in them (but not other things, such as coats or bags)
- Most accurate strategy took advantage of color profiles of empty seats versus those with people in them
- Iterations required to accommodate slight changes in camera angles
- Discrepancies reconciled during week of Daylight Savings change, when image time stamps diverged from actual times



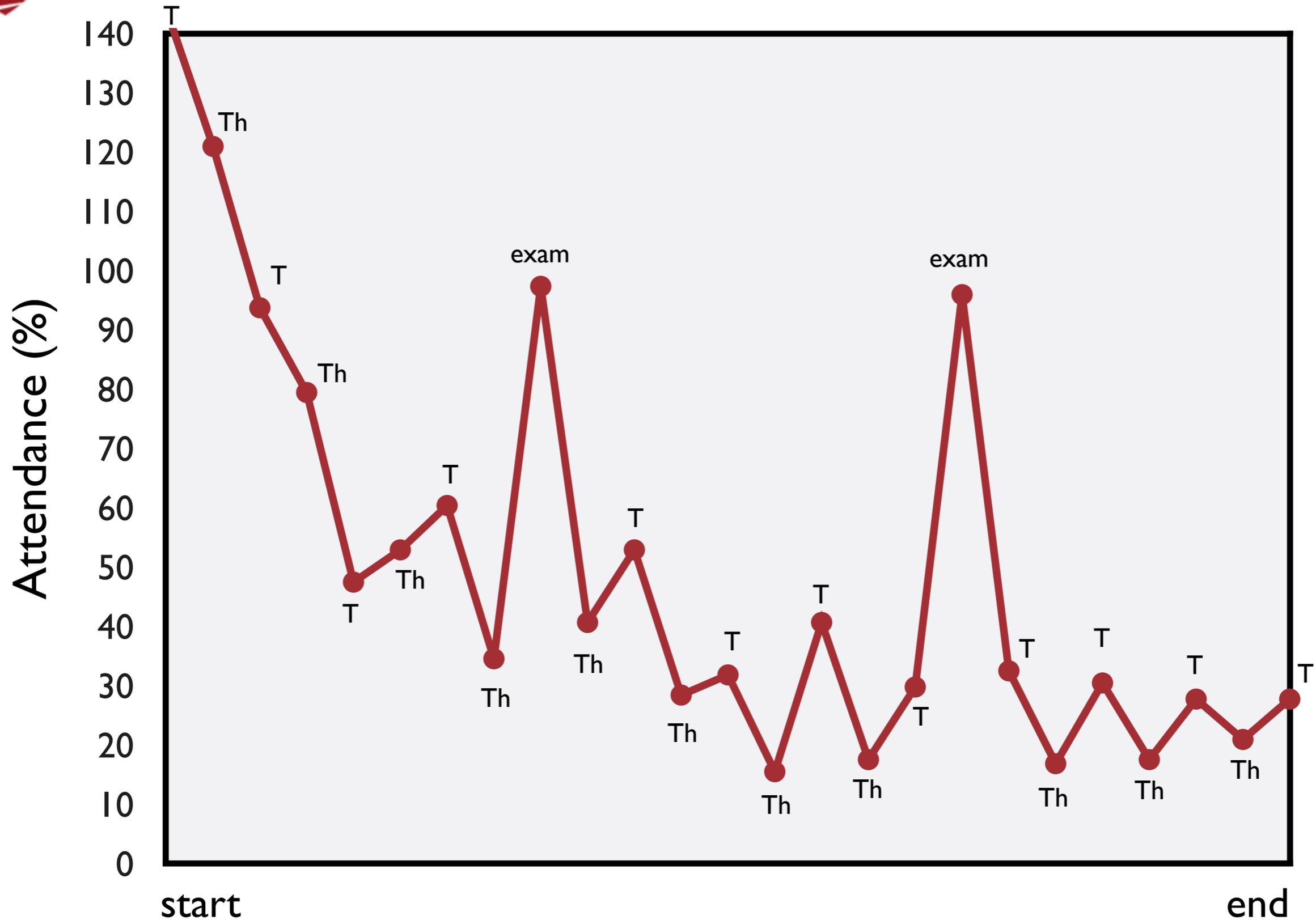
Validation of algorithm



- Algorithmic estimates compared with random sample of manually coded images
- Correlation extremely high ($r = .99$)
- Estimates plotted over time for each course, and potentially aberrant estimates identified
- Outliers scrutinized further based on input from faculty and inspection of imagery data



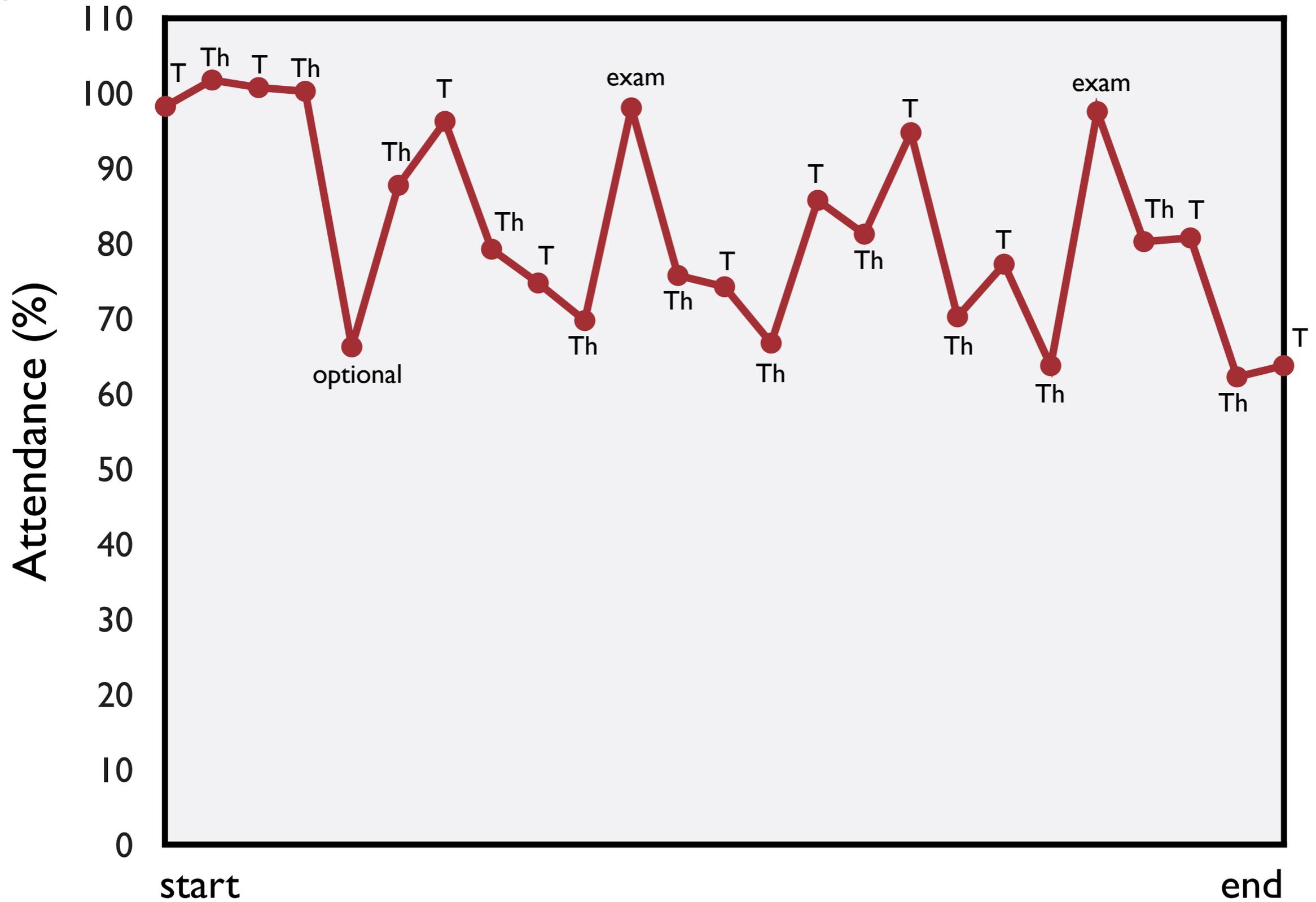
Course A (n ≈ 150)



At present time, attendance data from 10 courses have been analyzed fully and were presented at the HILT Conference 2014. Enrollment numbers and lecture dates were obscured for reasons of anonymity.

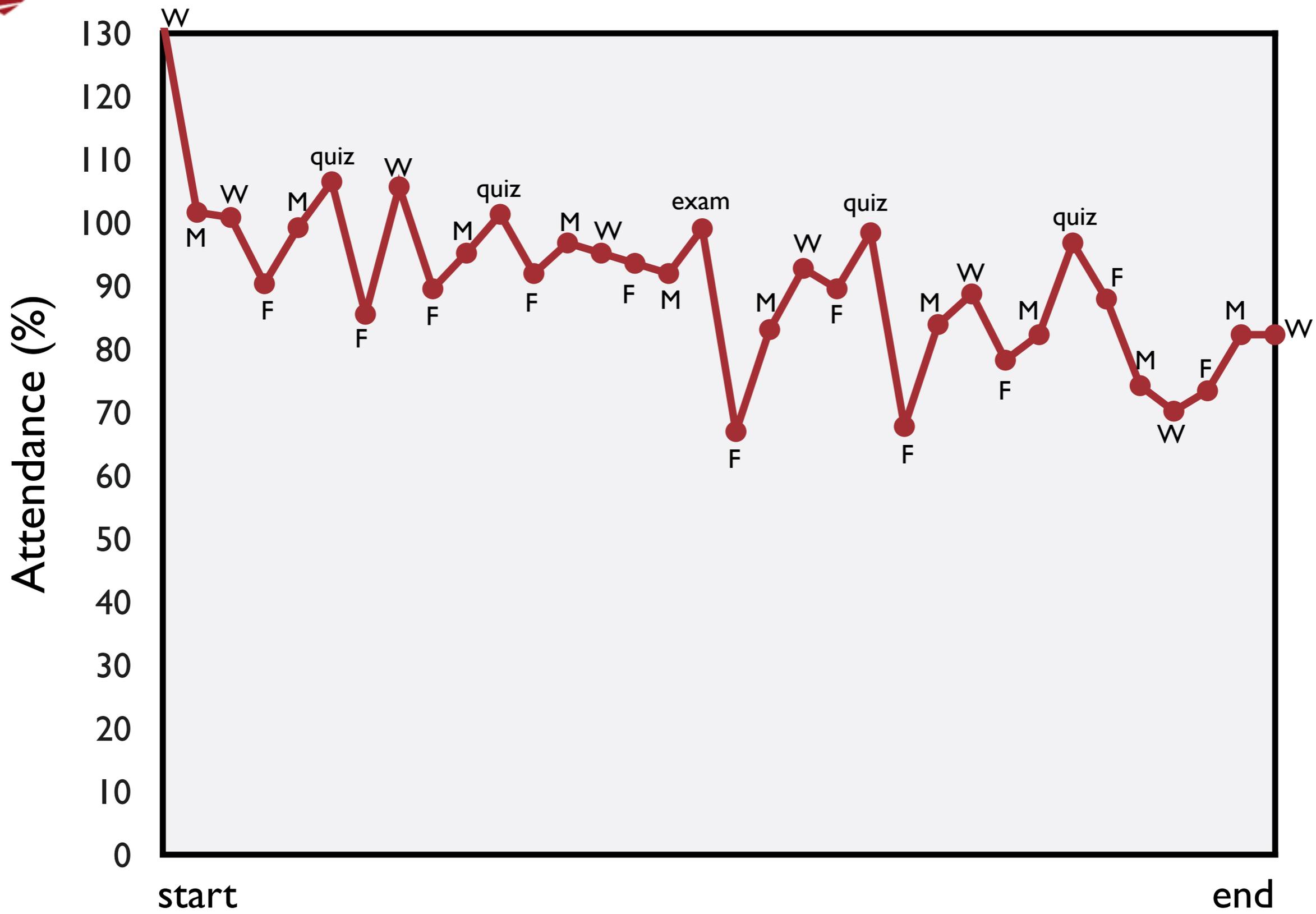


Course B (n≈200)



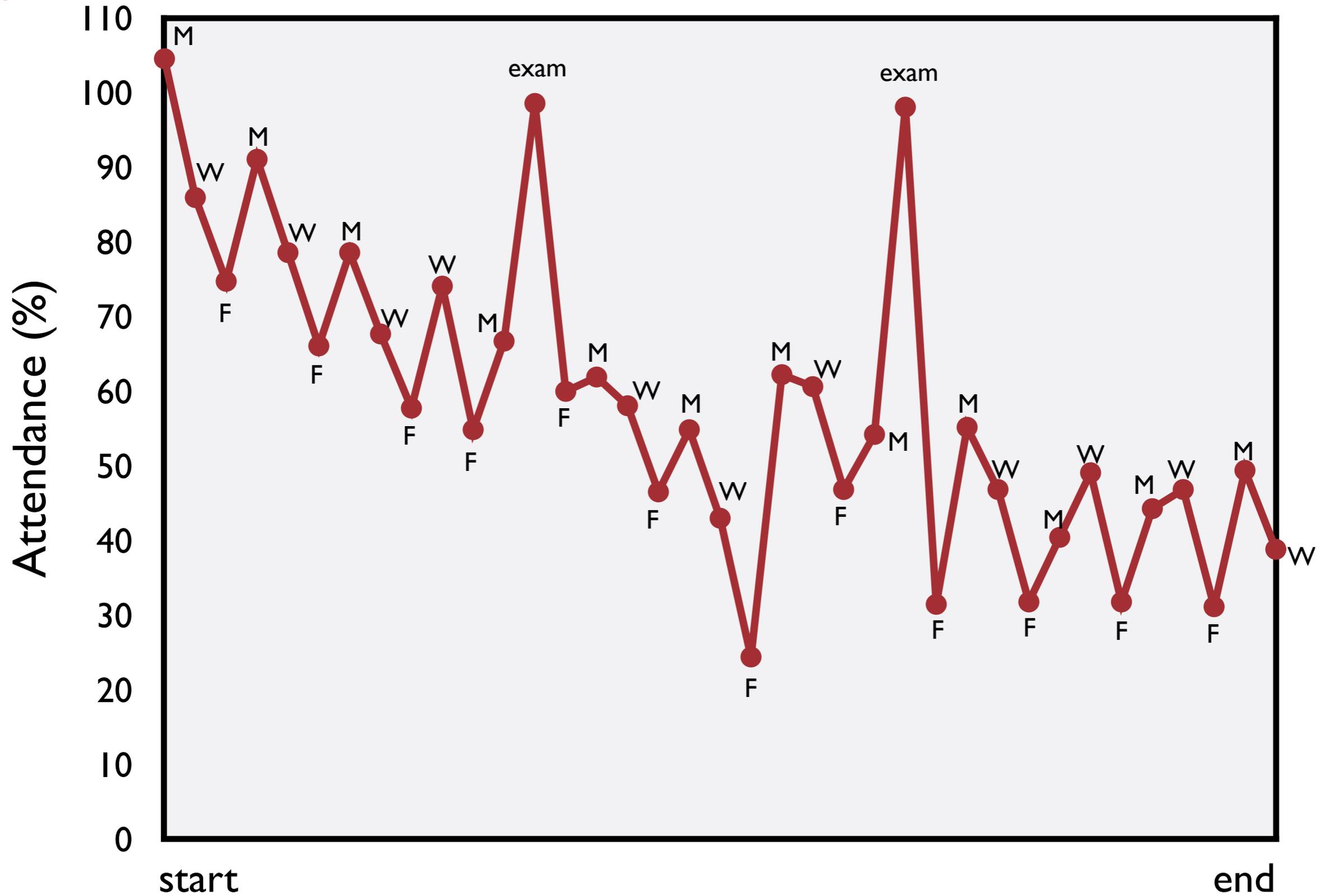


Course C (n ≈ 125)



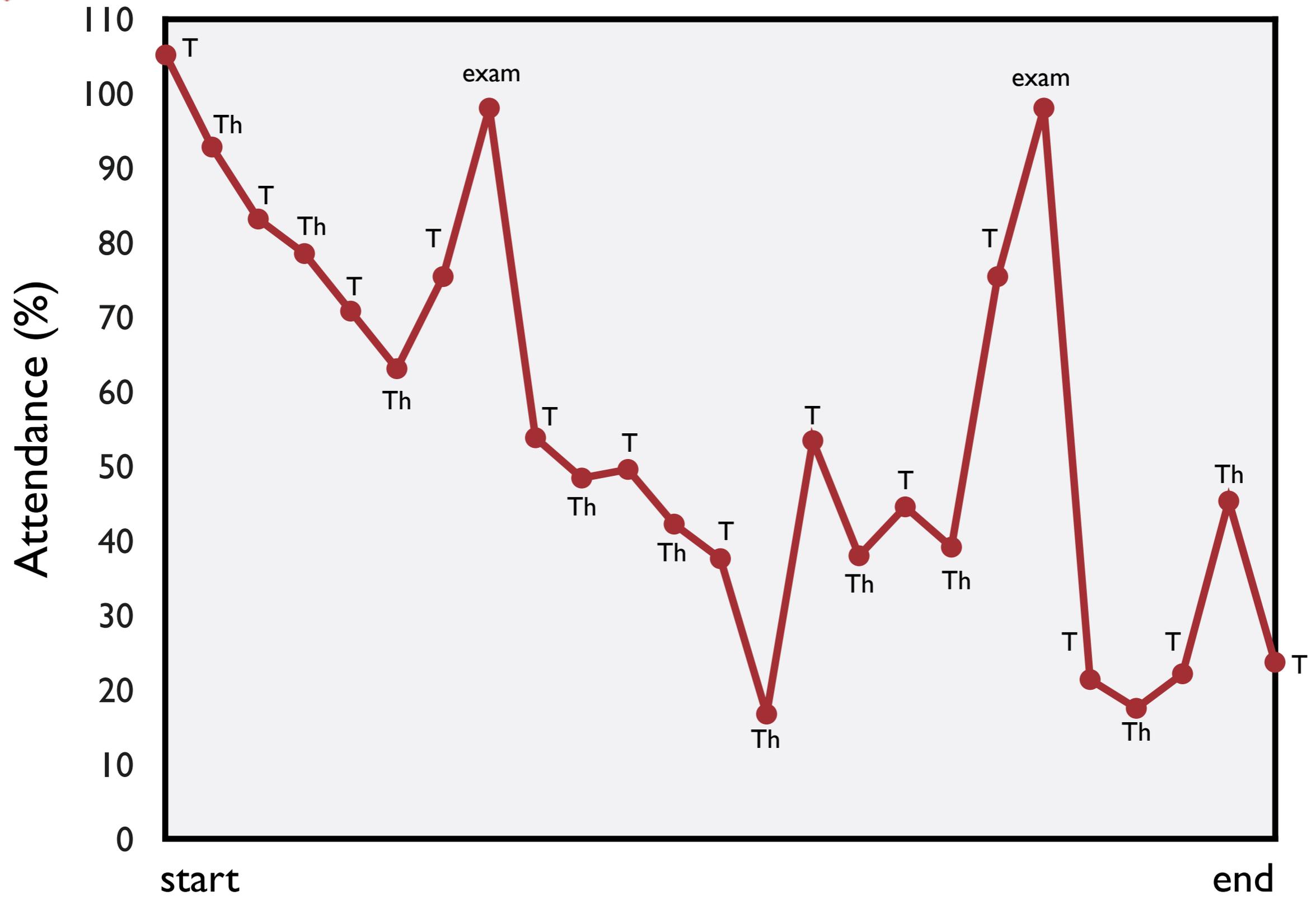


Course D (n≈300)



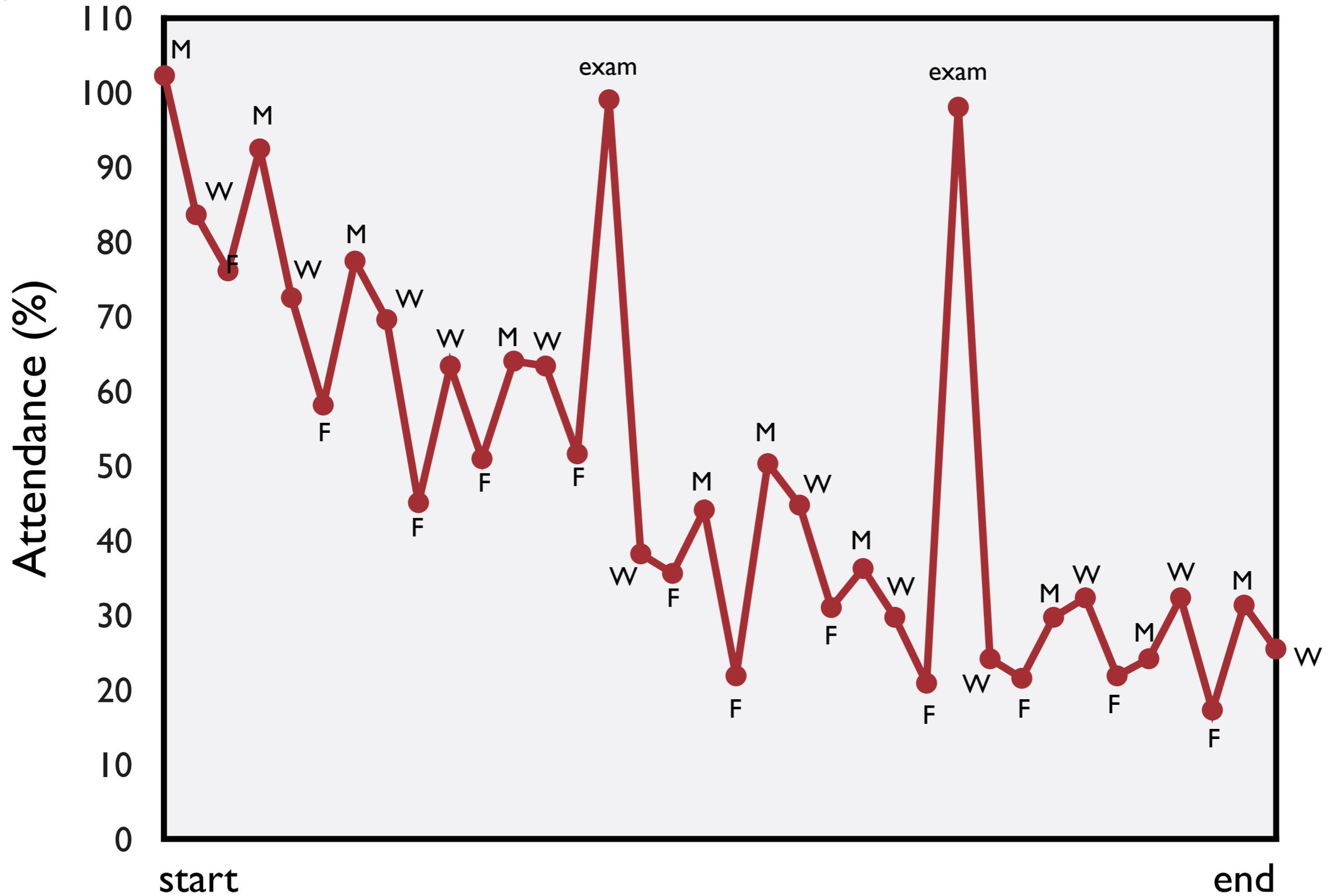


Course E (n≈250)



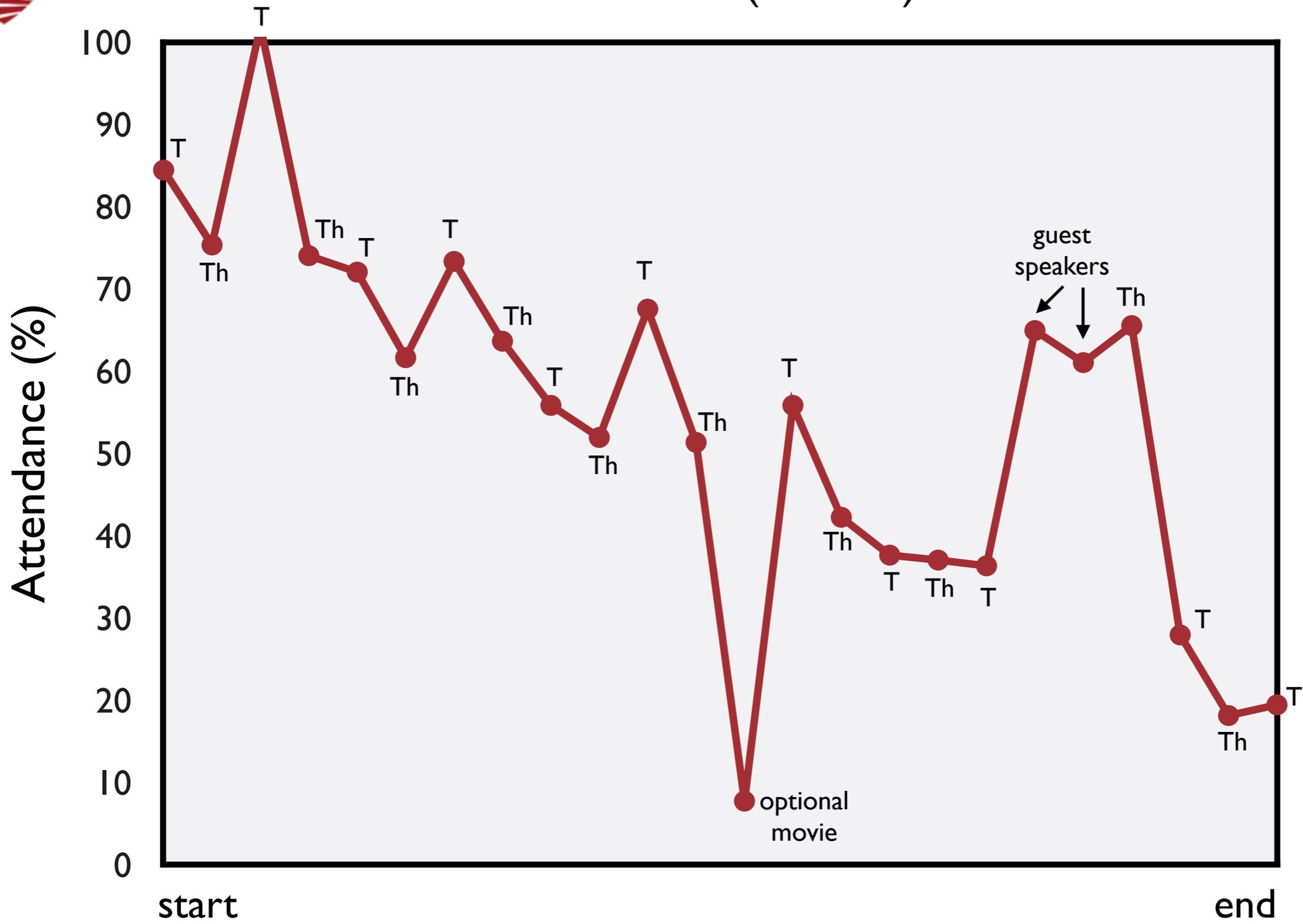


Course F (n≈300)



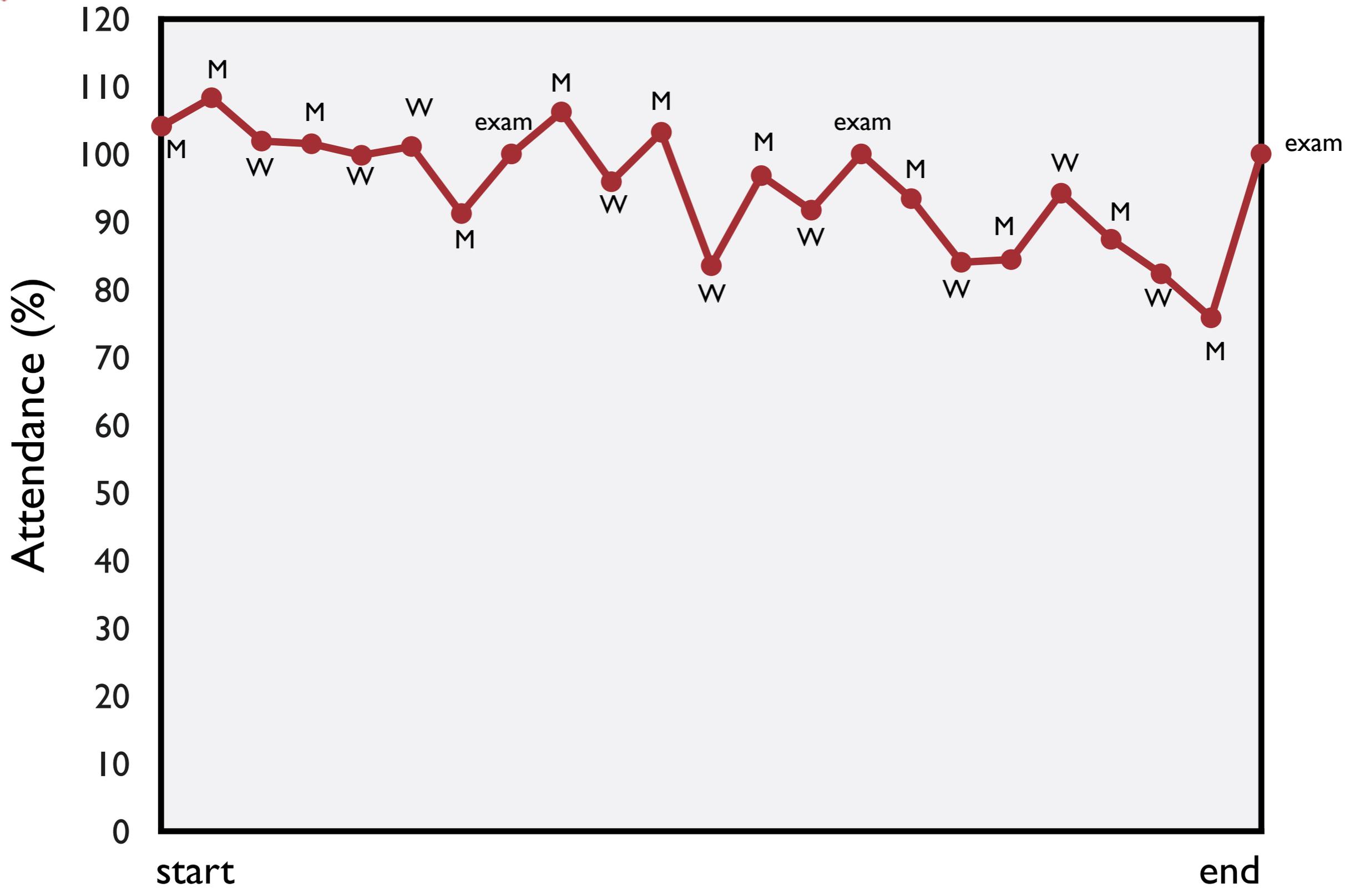


Course G (n ≈ 150)



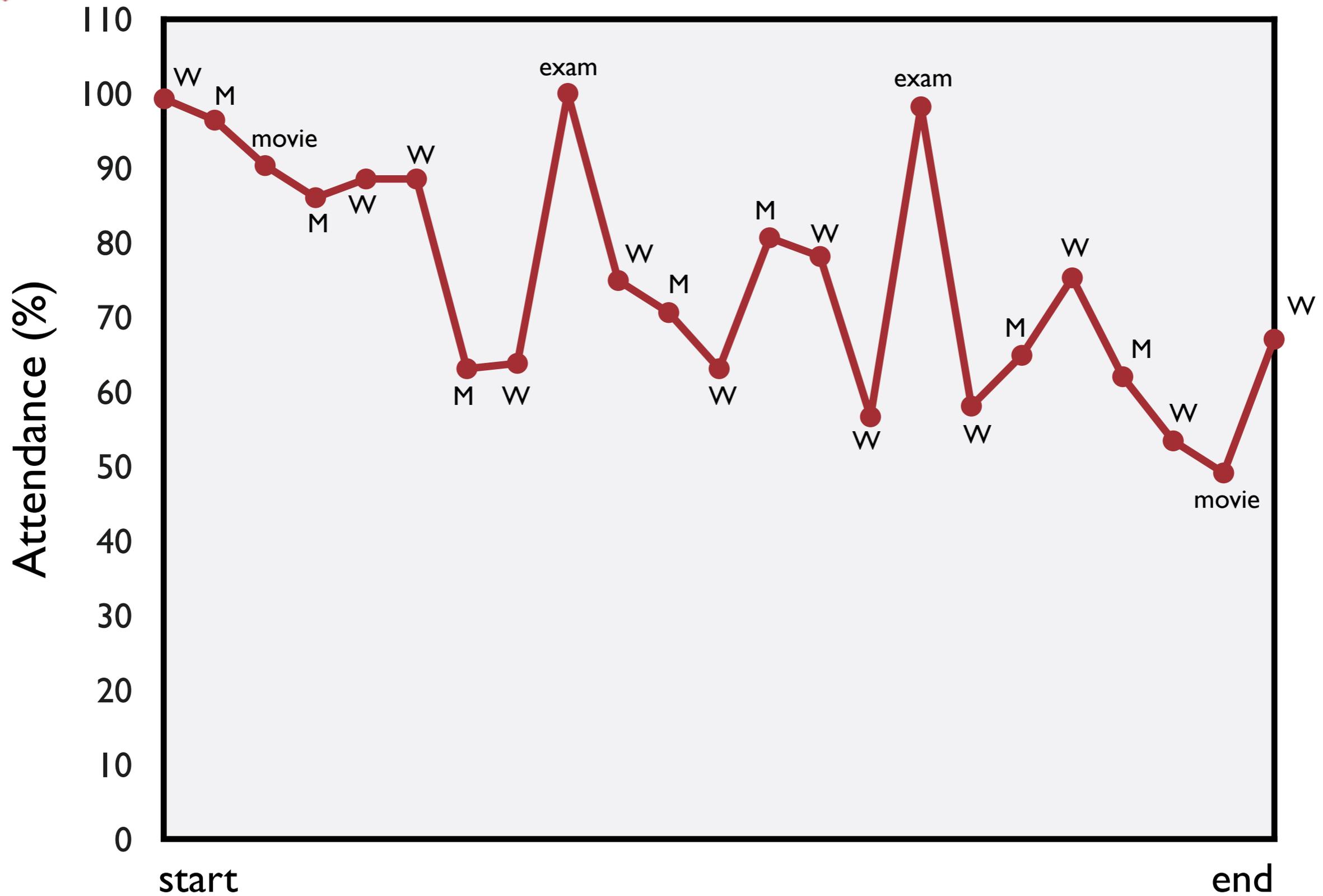


Course H (n≈250)



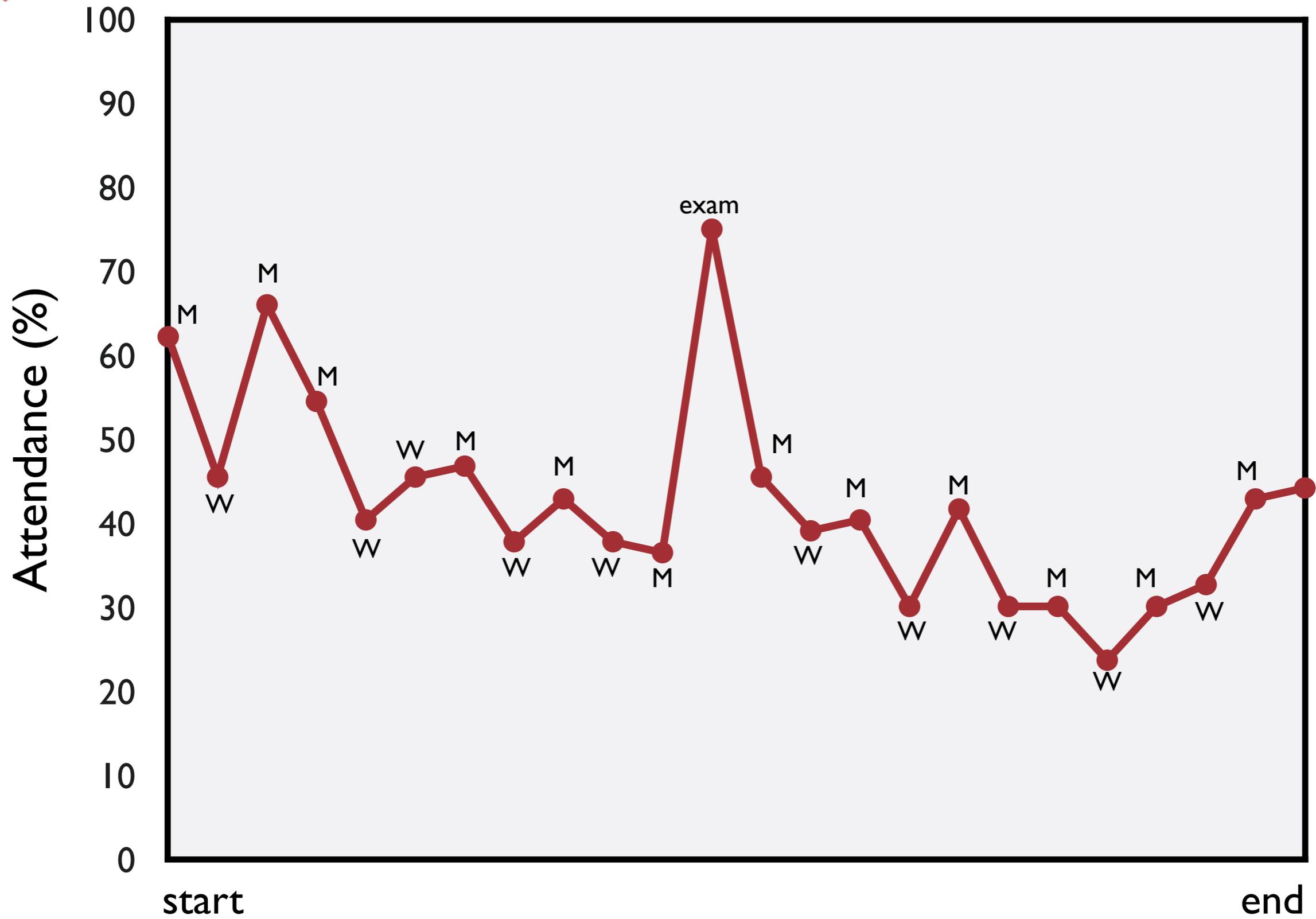


Course I (n ≈ 300)



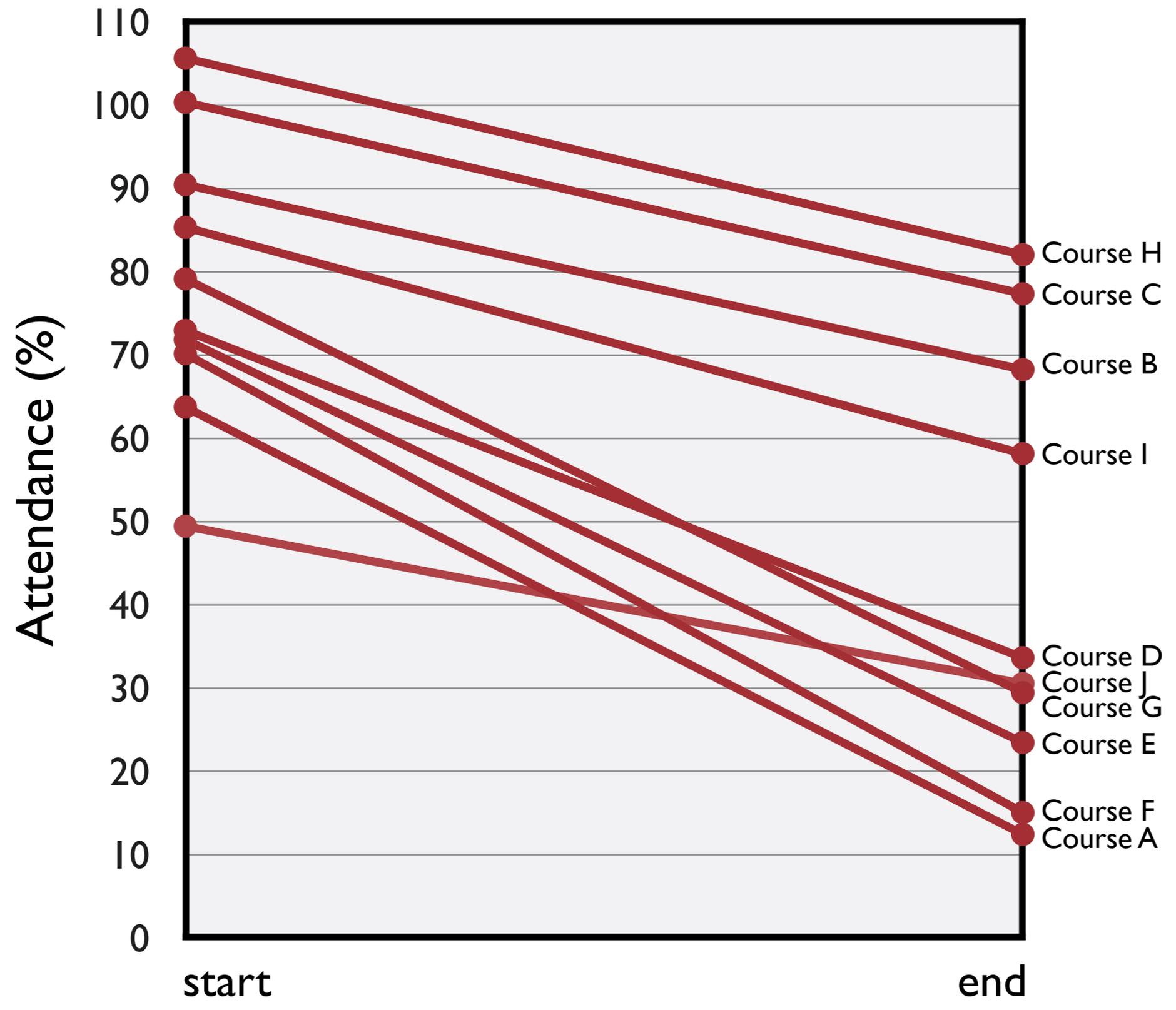


Course J (n ≈ 100)





Courses A-J





Courses A-J



On the average, 60% of students attended any given lecture.

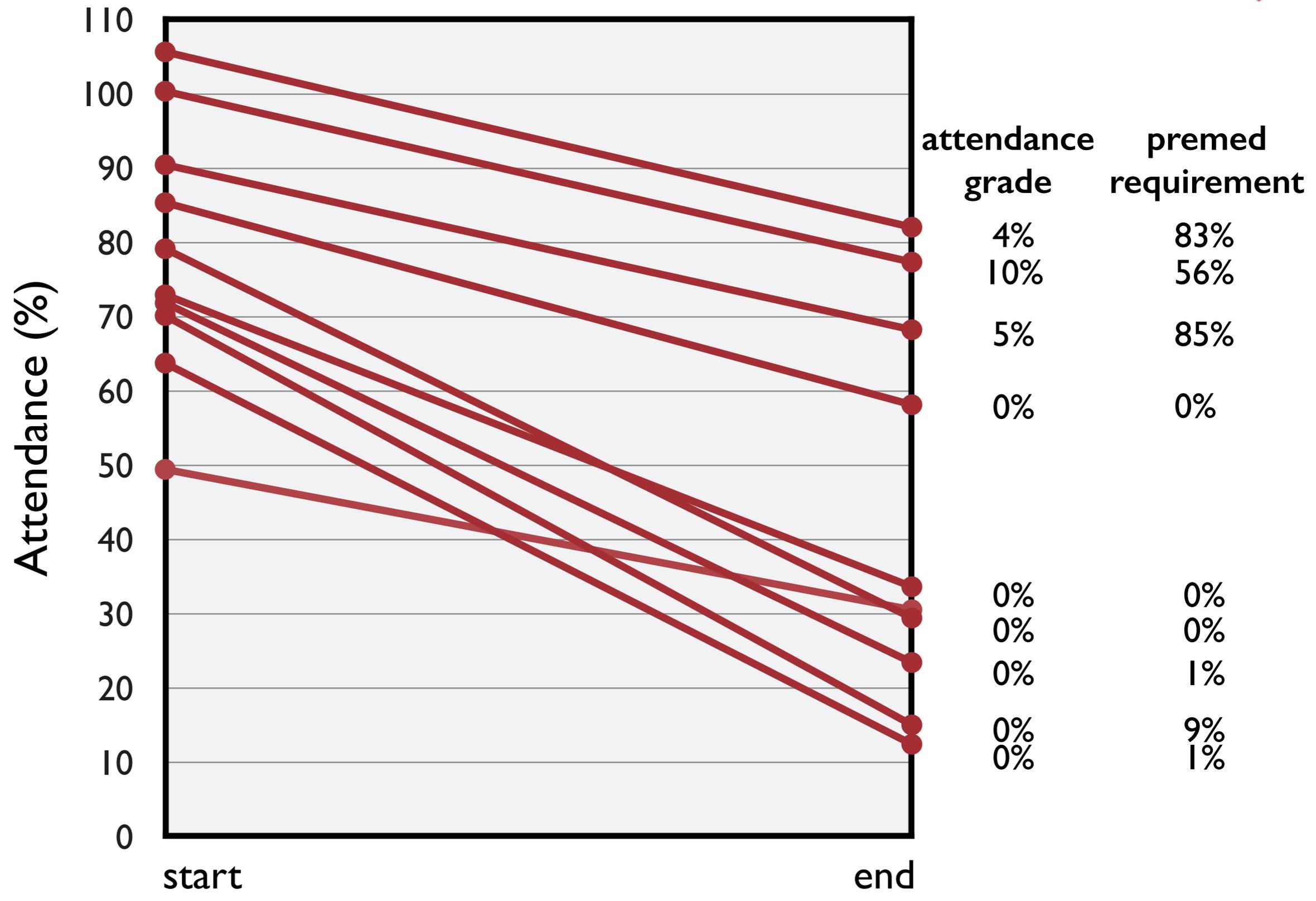
Moreover, there was incredible variability between courses in their average attendance: Some courses had as low as 38% attendance (over the semester), whereas others had as high as 94%.

Lecture attendance declined over the semester, starting at 79% and ending at 43%.

There was also loss of attendance over the week: 5% from Monday to Wednesday or Tuesday to Thursday, and 10% from Wednesday to Friday.



Courses A-J





Courses A-J



With this preliminary set of 10 courses, two noteworthy effects emerge.

First, courses that measured and graded attendance had higher attendance than those that did not (87% vs 49%, respectively). Second, courses in which students self-reported enrolling to fulfill a premed requirement had much higher attendance than other courses.

Other reasons for taking the courses (e.g., elective vs General Education requirement) did not show significant effects, nor did time of day, day of week, published Q ratings, or the availability of lecture videos.

Course I stands out as unusual: It had high attendance yet did not grade or measure attendance, nor did it fulfill a premed requirement. One distinguishing aspect of this course is that the instructor restricted access to the lecture videos until immediately before exams.



Preliminary findings

(10 courses)



Machine learning algorithms can be used to accurately measure attendance in large lecture courses.

Lecture attendance declines over the week and over the semester.

There is a huge range in average lecture attendance across courses.

These differences appear mostly structural in nature.



Site and video use



For most courses, site use (# unique authenticated users per week) correlated positively with lecture attendance. The more students attended lectures, the more they visited the course site.

Analysis of lecture video usage is problematic—most courses make their videos available to anyone at Harvard or anyone on the internet. Nonetheless, preliminary results show a positive correlation for some courses and a negative correlation for others.



Open questions



How do we interpret attendance in high-enrollment courses?

What evidence-based advice can we offer instructors who wish to promote attendance in high-enrollment courses?

How does the availability and use of online course videos (or other digital resources) affect attendance in high-enrollment courses?

Will the present findings replicate in a larger sample of high-enrollment courses? What other factors explain variability between courses in lecture attendance?