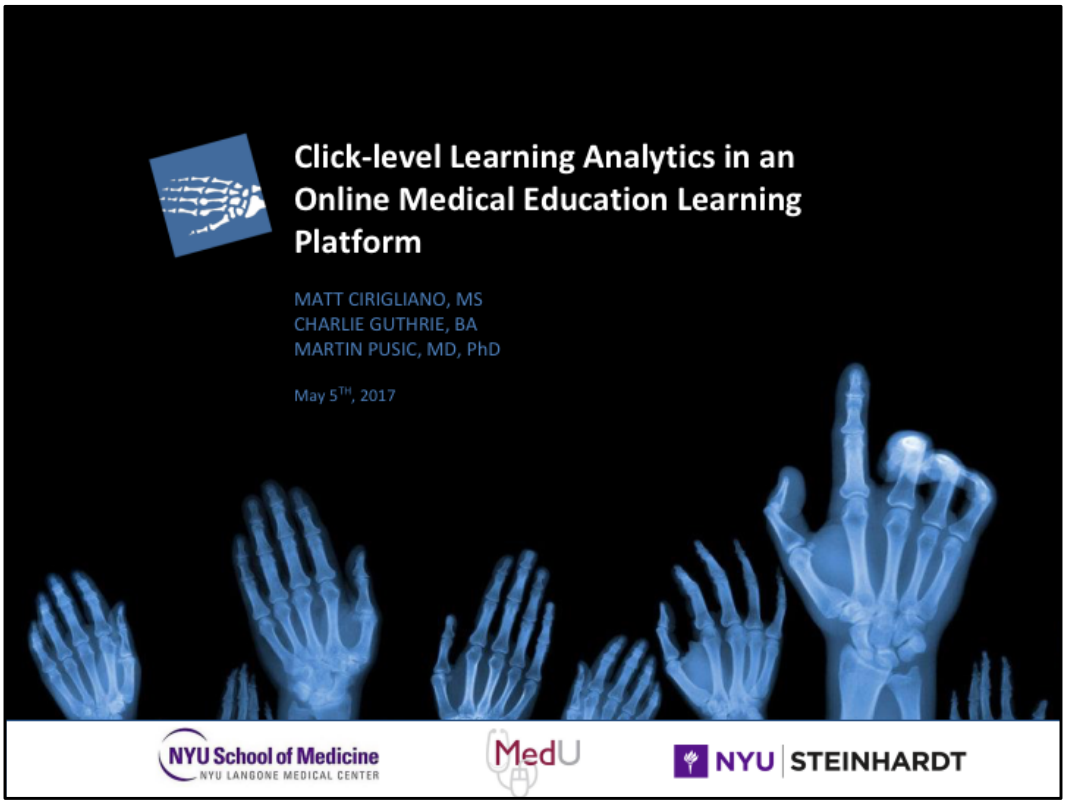


Hi everyone, my name is Matt Cirigliano and I am a **Doctoral Candidate at NYU Steinhardt** working on research in medical education and the learning sciences. With me I have…

Charlie Guthrie - I just finished my Master's in Data Science at NYU, with the Center for Data Science. My background is in statistics with a focus on learning analytics.

And we're currently working with **Dr. Martin Pusic** -- **who [you've just met in the last talk] is the Director of the Division of Learning Analytics** at the **Institute for Innovations in Medical Education at the NYU School of Medicine** -- on click-level learning analytics in MedU, an online medical education learning platform.

[26 sec]

Our objective was to understand **measures of learner engagement**…

…meaning **what learners clicked on**, interacted with, and for how long…

…and **how these behaviors related to learner achievement**. MedU's historical database of learner interactions and **learning analytics** allowed us to do that.

[14 sec]

LEARNING ANALYTICS

…is an emerging field in which analytic tools adapted from computer science, math, and statistics are used to improve learning and education by extracting usable information from very large datasets.

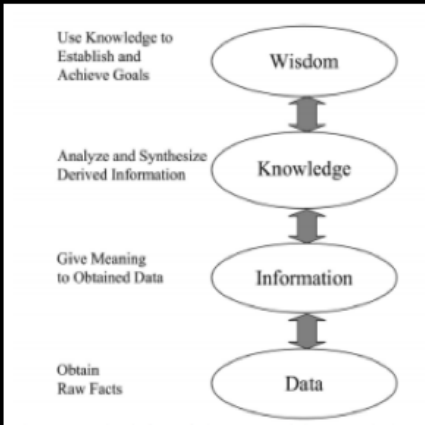


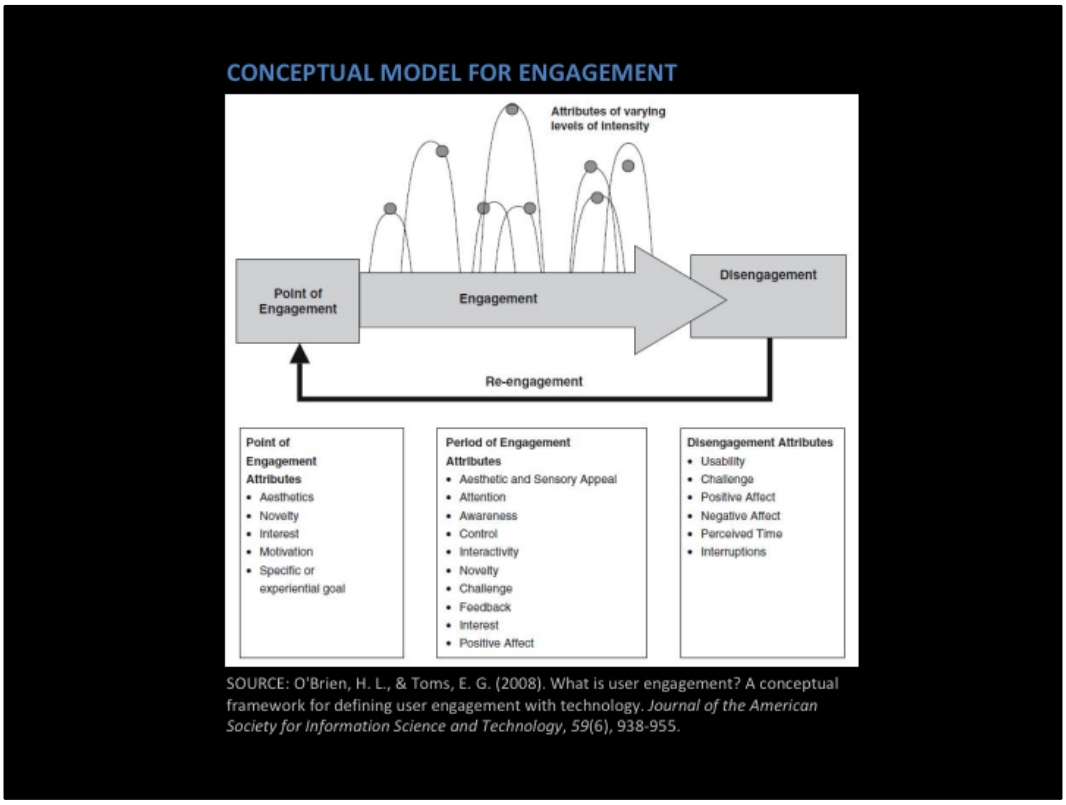
Figure 1: Baker's (2007) depiction of the Knowledge Continuum.

SOURCE: Elias, T. 2011. *Learning Analytics: Definitions, Processes and Potential.*

CORE RADIOLOGY COURSE

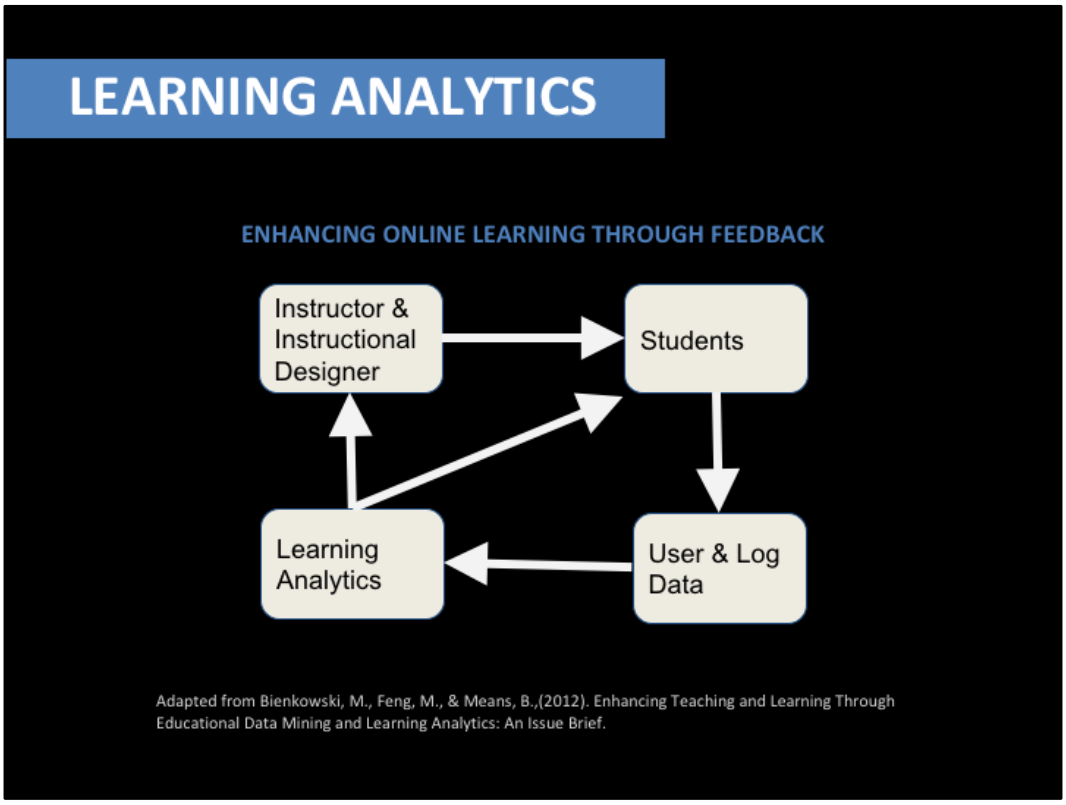
Briefly, learning analytics uses the **power of large datasets and analytic tools** to understand how learners engage with material and improve approaches to achieving **educational goals**.

[11 sec]



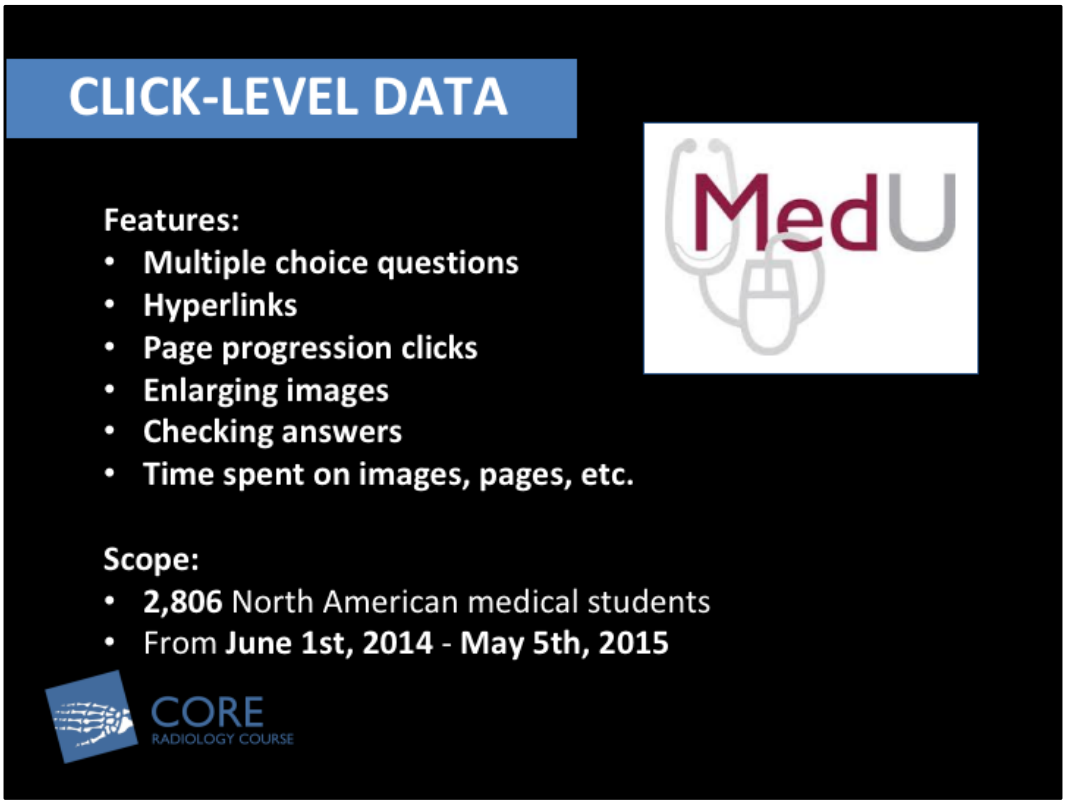
And engagement incorporates the **complex network of interactions a learner** has with content—if they become engaged or disengaged by material. **Learning analytics** and **predictive models** can help us identify what content is most/least useful.

[14 sec]



Overall, learning analytics can help us **generate feedback systems to help** stakeholders improve learning content and strategies.

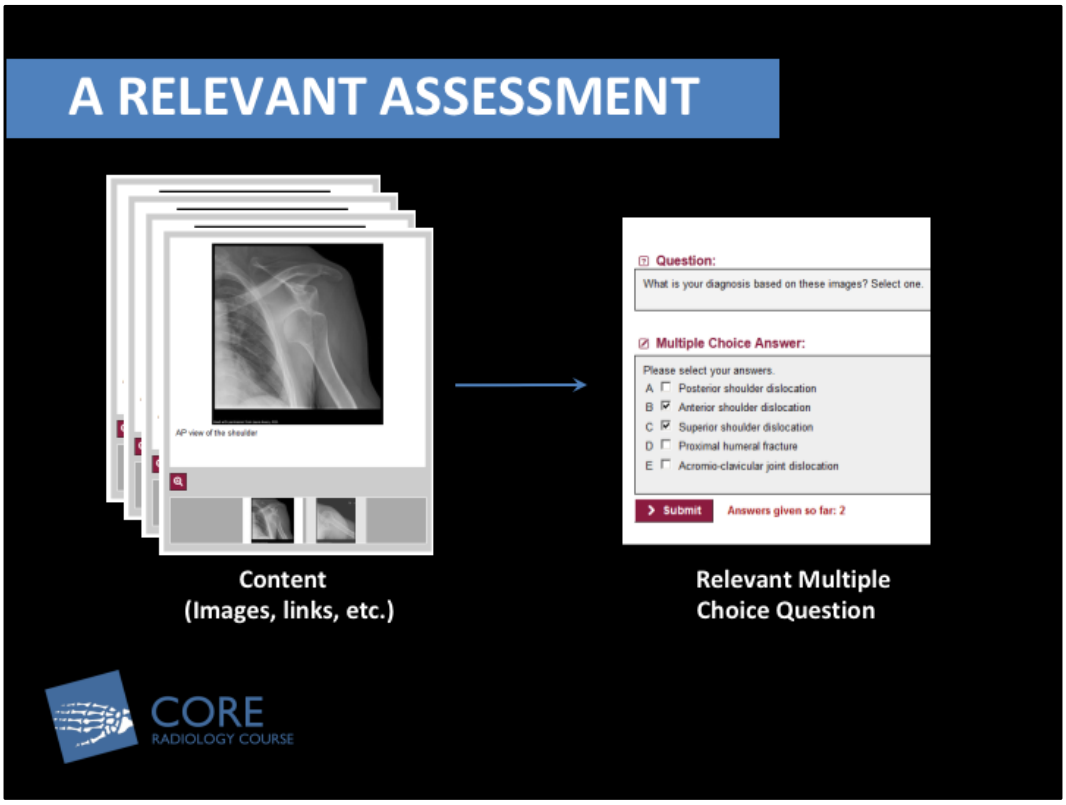
[9 sec]



So with Med-U, we applied learning analytics to click-level data to reveal how learners interacted with the content. That includes [riff]

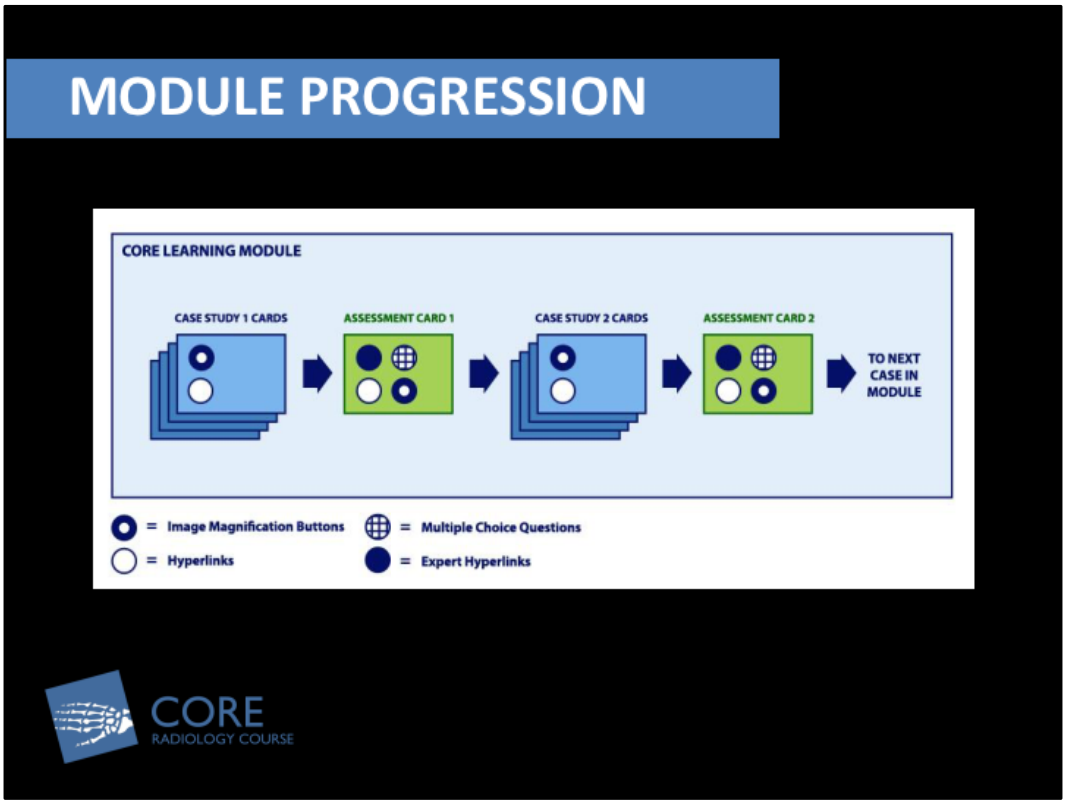
MedU itself is an online suite of case-based learning systems and courses accessed by over 150 different medical schools across North America—you can see some more information at the bottom—there were over 2800 med students who contributed (only 6). One feature of MedU is the CORE Radiology series, which has 18 modules total. We focused on one on musculoskeletal trauma.

[32 seconds]



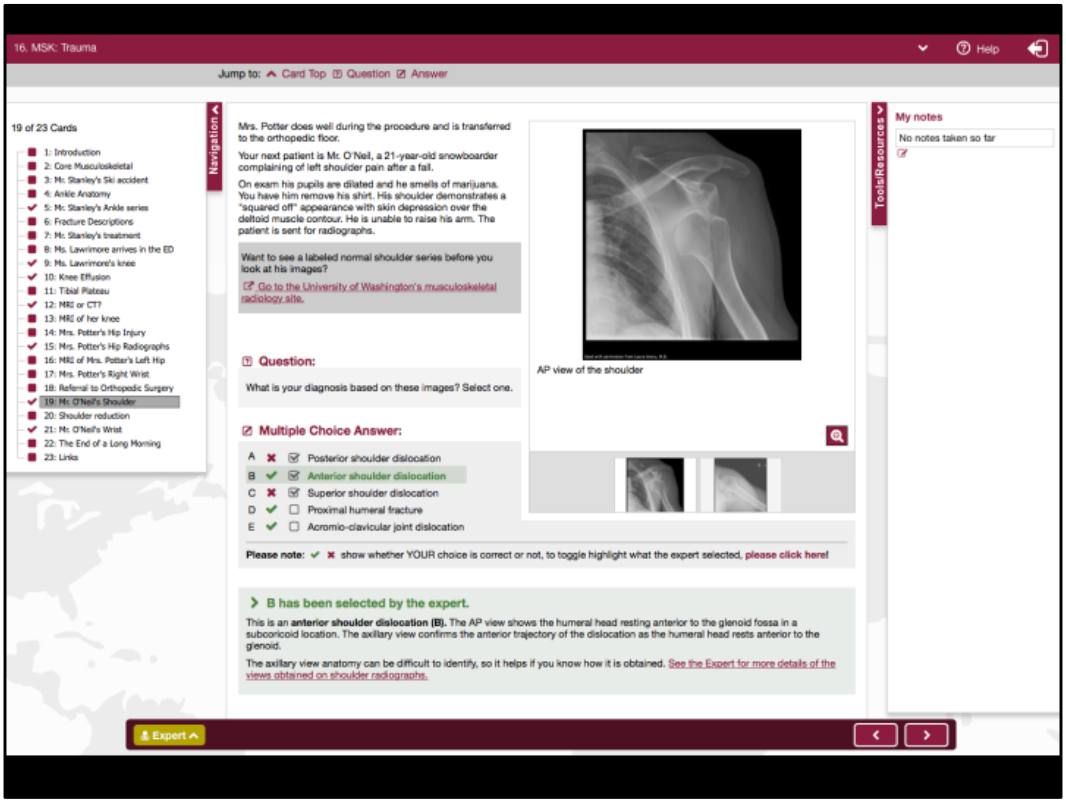
So in sum, we wanted to understand if engaging with relevant content impacted performance on assessment questions in the module.

[7 sec]



The module was **broken down into units,** where a set of **cards** featuring content was followed by an **an assessment card**, which featured a relevant multiple choice question.

[9 sec]



And the whole module was **23 cards long**. This is a **screenshot of one card**.

[5 sec]

But...which features would be worth exploring with learning analytics?

[3 sec]

FOCUS GROUP RESULTS

Six experts grouped and ordered candidate analytic measures (CAMs), revealing which were considered the most useful:

(1) Thumbnail Click
(2) Post-Answer Expert Feedback Use
(3) Supplementary Link Click
(4) Magnifying (Zooming-in on) Images
(5) Time Spent on Cases

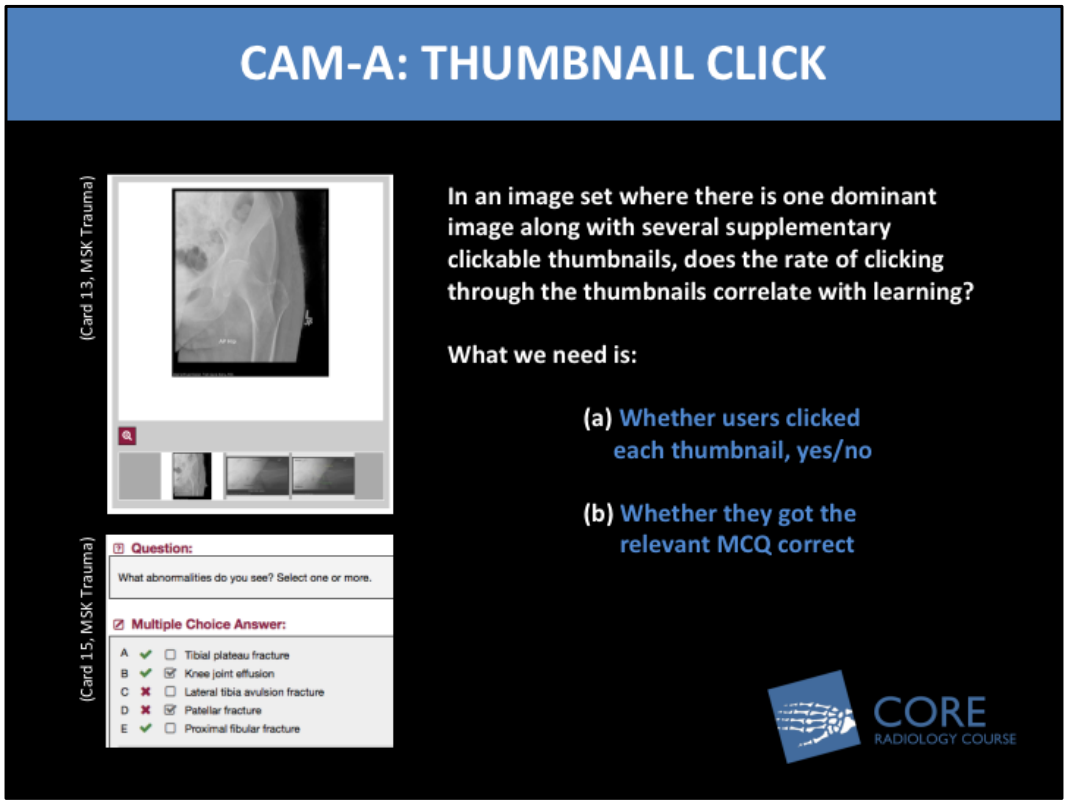
Ranking of Candidate Learning Analytics Measures (CAMs)

	Candidate Learning Analytic Measures					
	1 Thumbnail Click	2 Supplementary Link Click	3 Duration on Cases/Images	4 Zooming-in on Images	5 Post-answer Feedback Use	6 Proportion of Feedback Used
Expert 1	1	12	8	2	5	11
Expert 2	1	2	8	5	3	4
Expert 3	1	2	1	2	1	1
Expert 4	4	2	6	5	3	1
Expert 5	2	2	5	2	4	4
Expert 6	4	4	4	8	1	8
TOTAL	**13**	**24**	**32**	**24**	**17**	**29**
Std. Dev.	**1.5**	**4.0**	**2.7**	**2.4**	**1.6**	**4.0**

	Candidate Learning Analytic Measures					
	7 Labeled Peer Answers	8 Viewing Sequence I	9 Relative Time on Views	10 Video Playback Speed	11 Opposing Design Choices	12 Forced Views
Expert 1	6	4	10	9	3	7
Expert 2	9	6	7	5.8	6.4	6
Expert 3	3	3	3	3	3	1
Expert 4	7	8	9	10	11	12
Expert 5	1	3	5	6	7	2
Expert 6	1	4	8	1	8	8
TOTAL	**27**	**28**	**42**	**35**	**38**	**36**
Std. Dev.	**3.3**	**2.0**	**2.6**	**3.4**	**3.1**	**4.0**

To identify **Candidate Analytic Measures,** we performed a focus group with experts in medicine and instructional design to see what they thought would be most important to know about learner behavior. **A set of 12 analytics were ranked** and the top five were selected for further study.

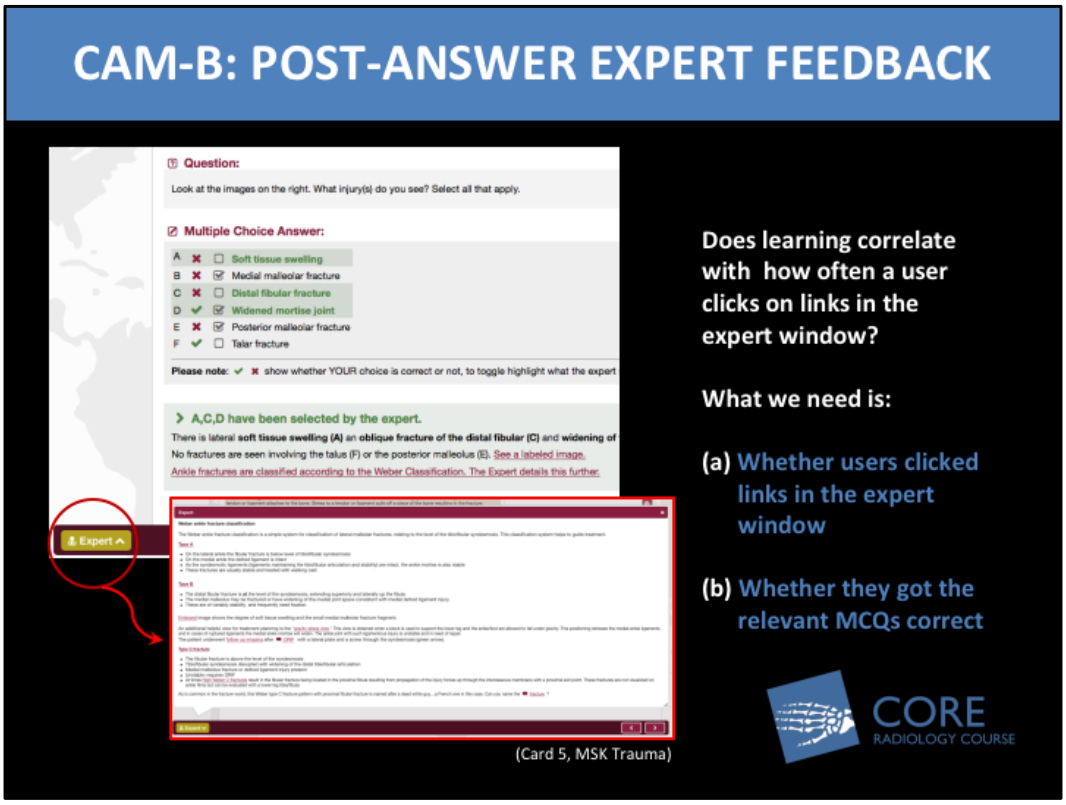
[18 seconds]



The first was the thumbnail click, and whether clicking on relevant thumbnails impacted assessment outcomes.

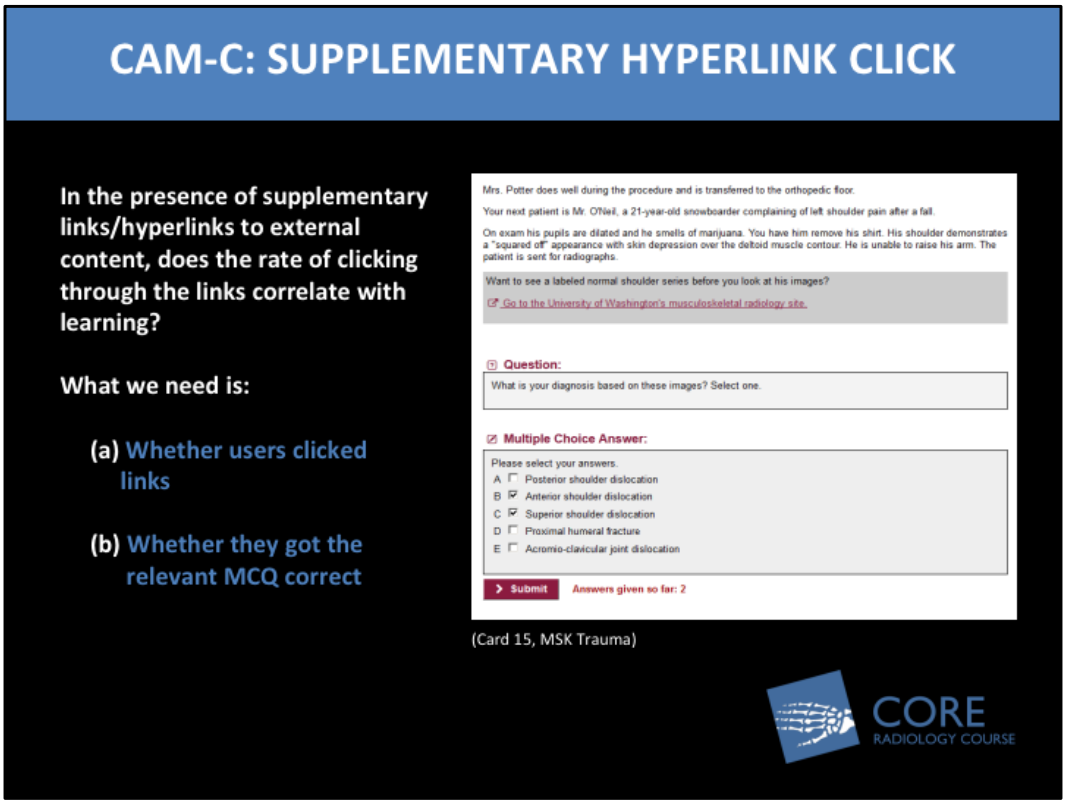
Sadly, because this data was unavailable in the database, it wasn't incorporated in the model. But there's more…

[11 sec]



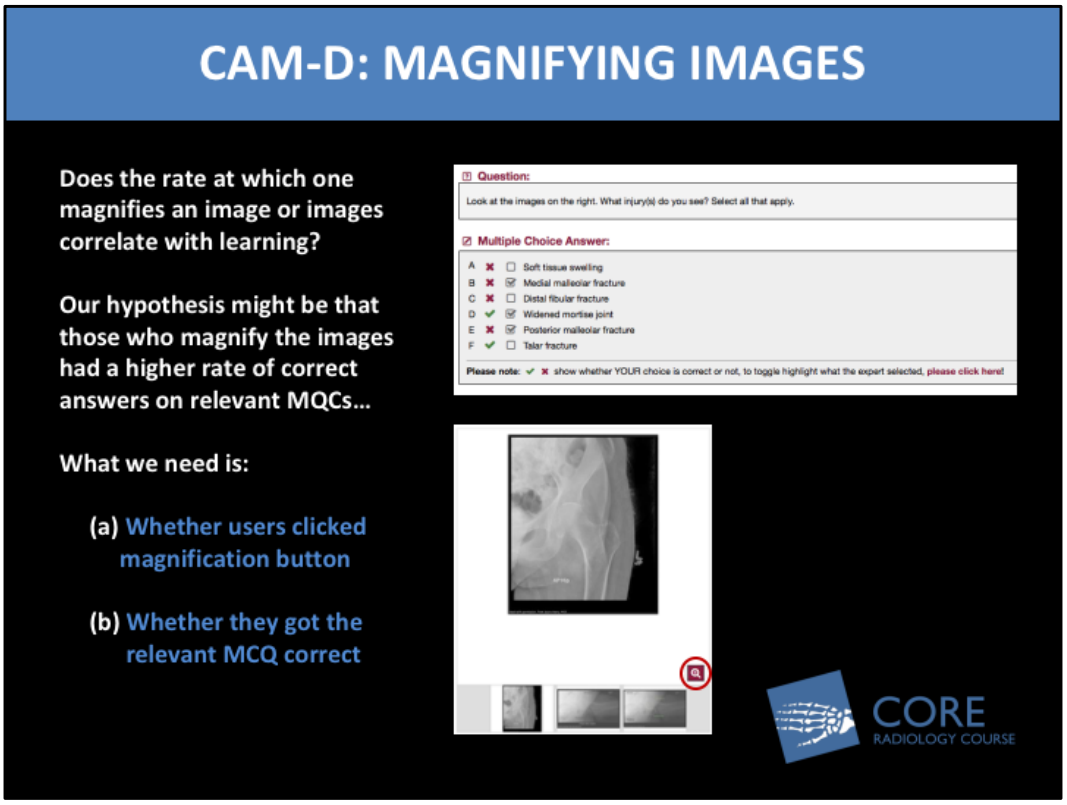
The next measure is clicking on “expert links”, which showed how experts might respond to questions.

[6 sec]



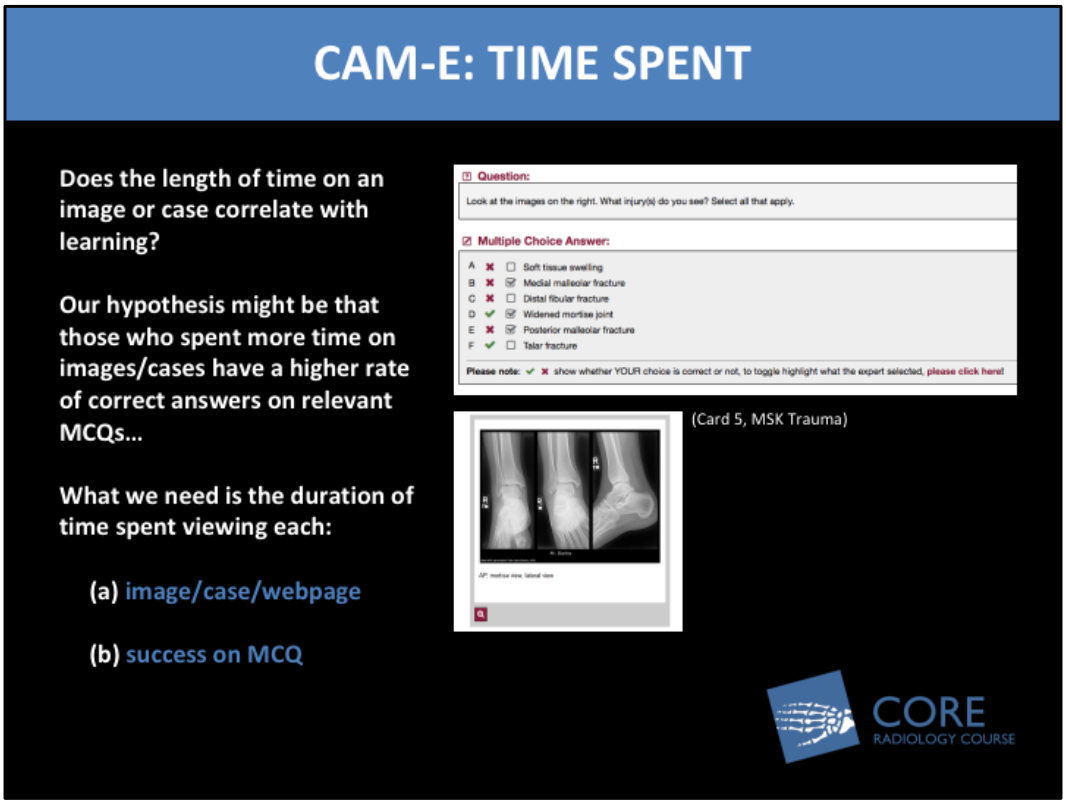
Third was clicking on hyperlinks, and whether this predicted better assessment outcomes.

[5 sec]



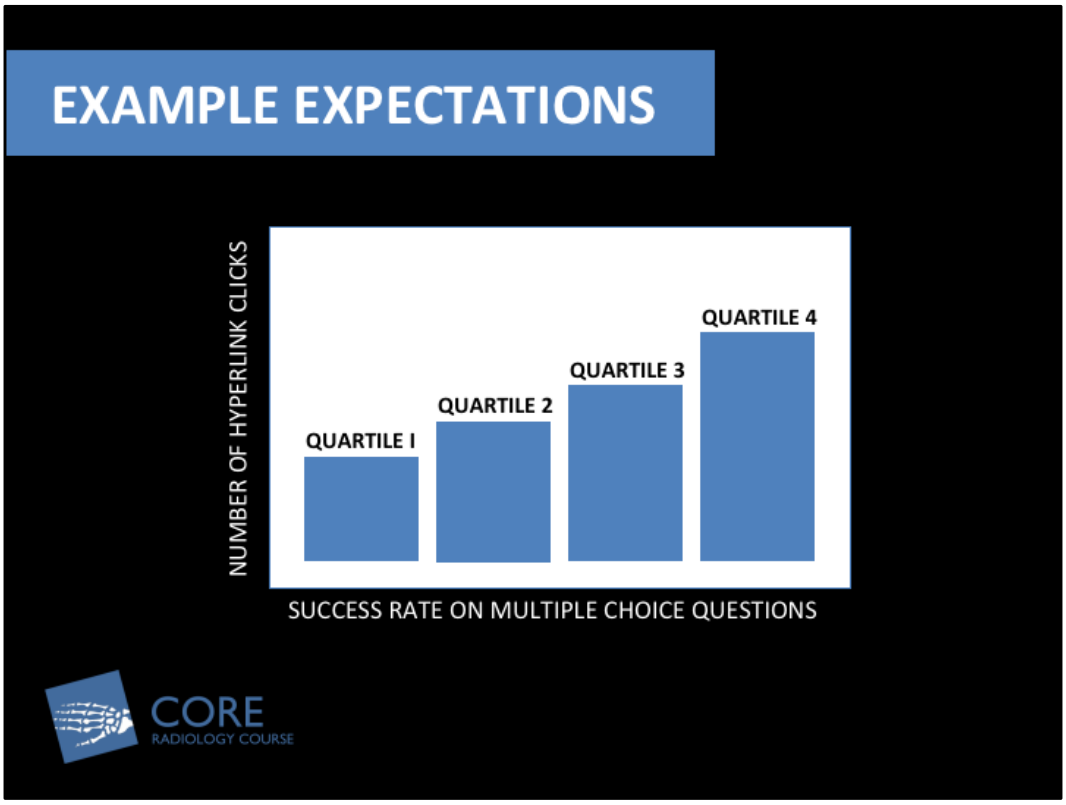
Fourth was magnifying or zooming in on images, and whether this behavior predicted better outcomes.

[6 sec]



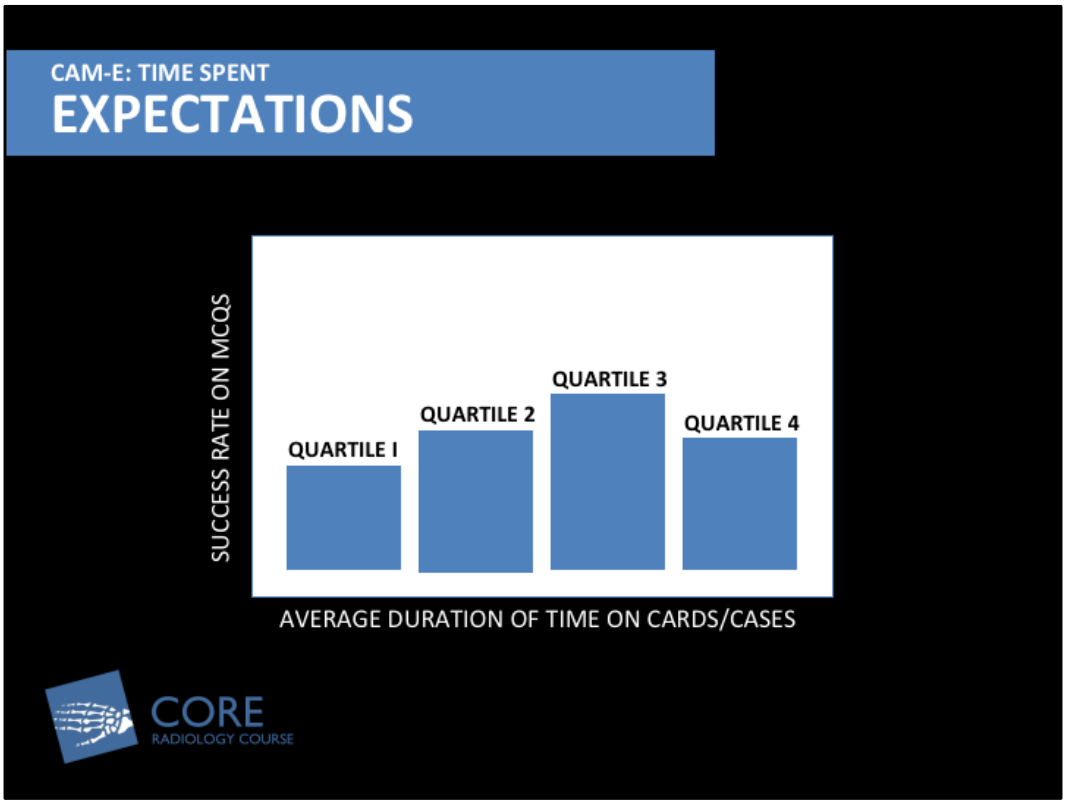
And finally, the fifth measure was time spent on each card, and whether this behavior had a relationship with performance.

[7 sec]



In terms of expectations, we might expect that more clicks on links and images would predict better outcomes, since those learners engaged with more content.

[9 sec]

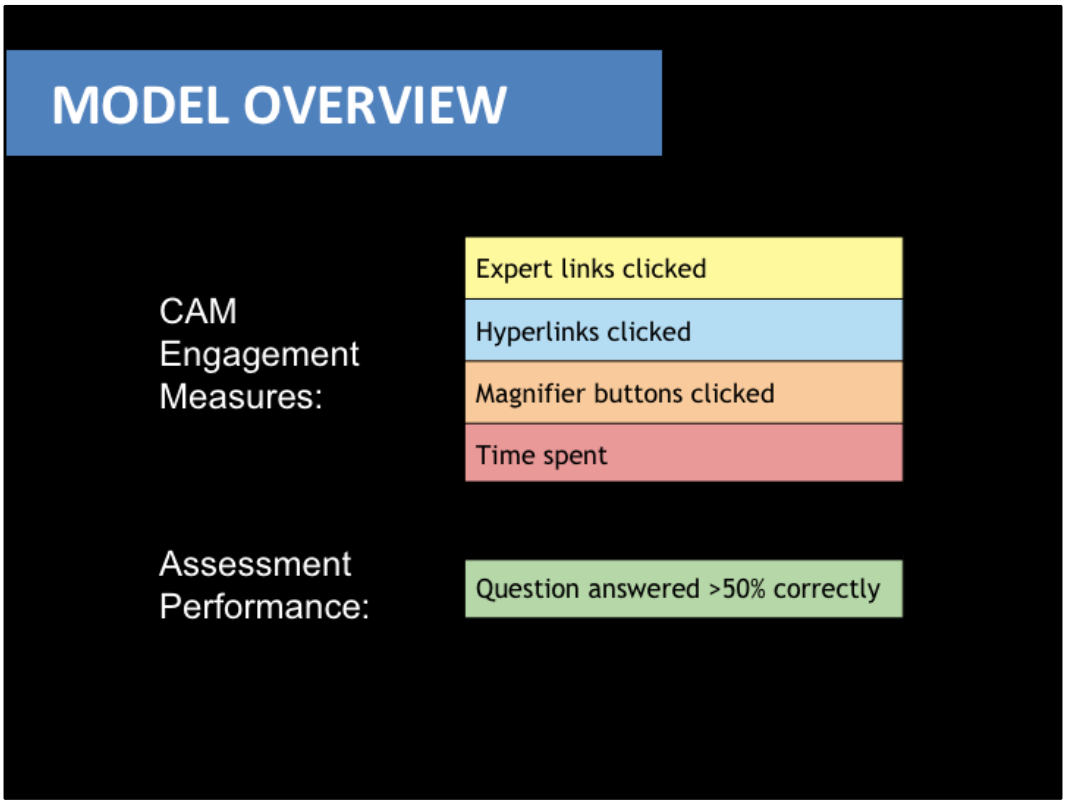


And more time spent on a card would also be expected to result in better assessment outcomes, with the exception of very long times, as these might indicate off-task behavior.

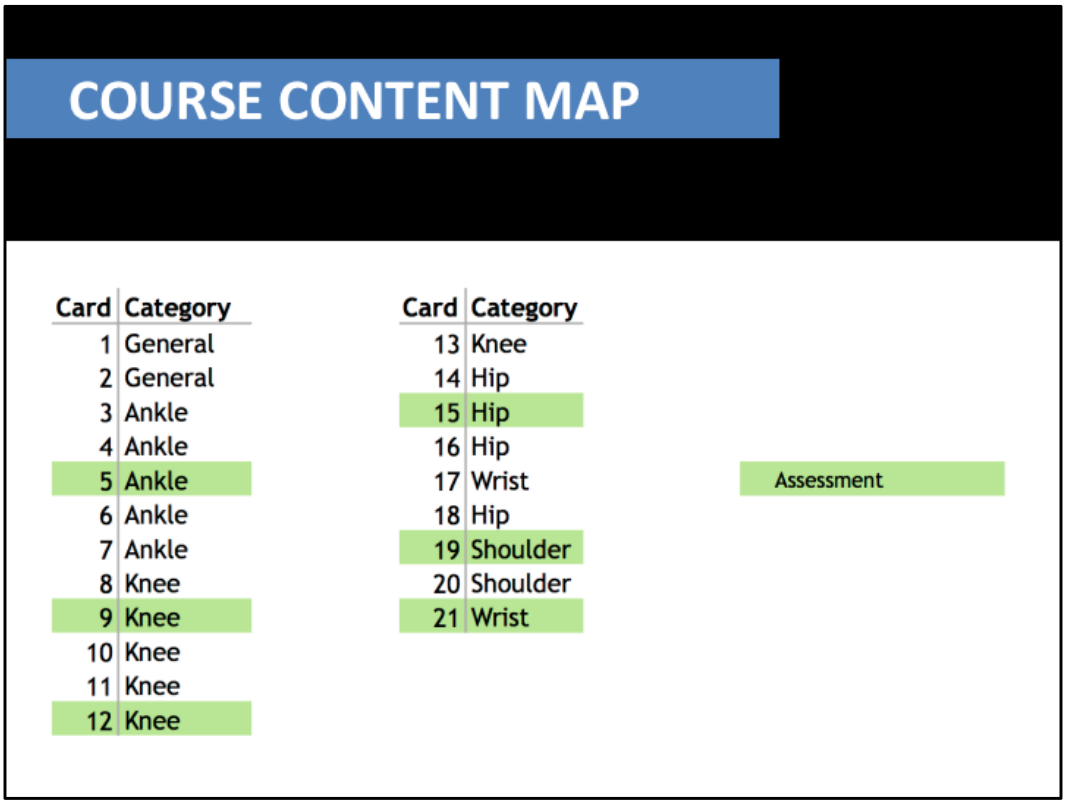
So, what did we find? *[Hand off to Charlie]*

[15 sec]

Having hypothesized about which activities would correlate with assessment performance,
we set out to build models to test them.

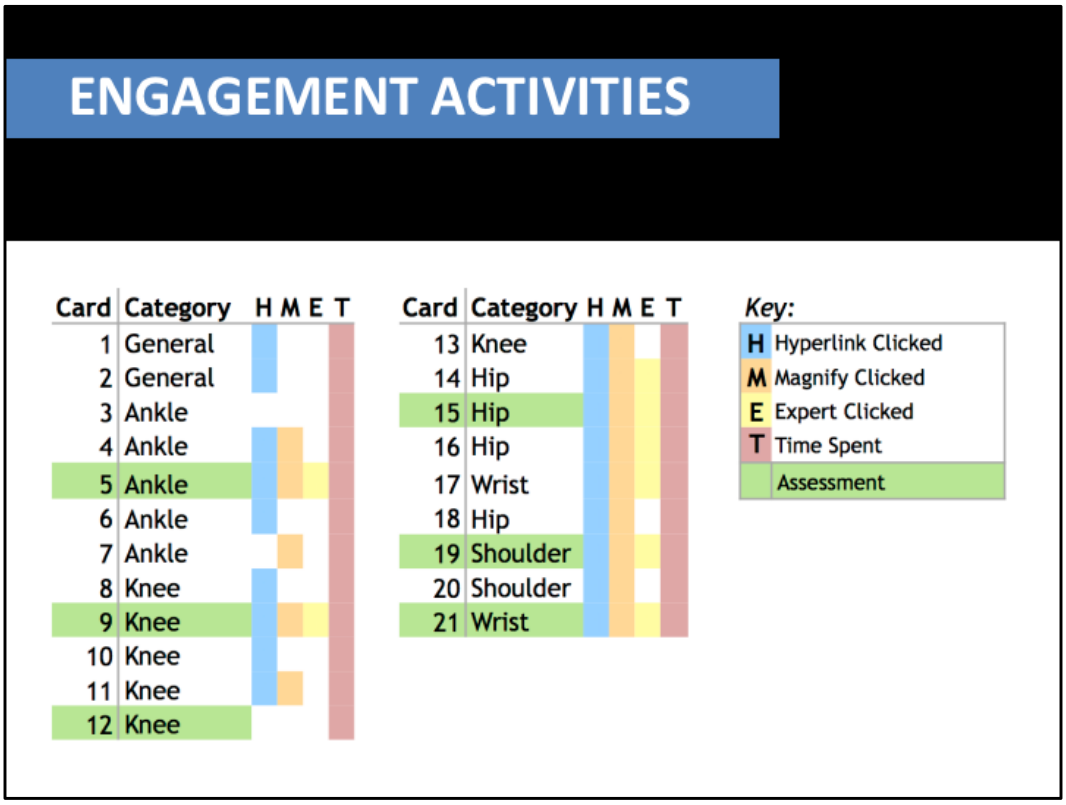


There were two investigations. Both involved using engagement measures to predict assessment performance

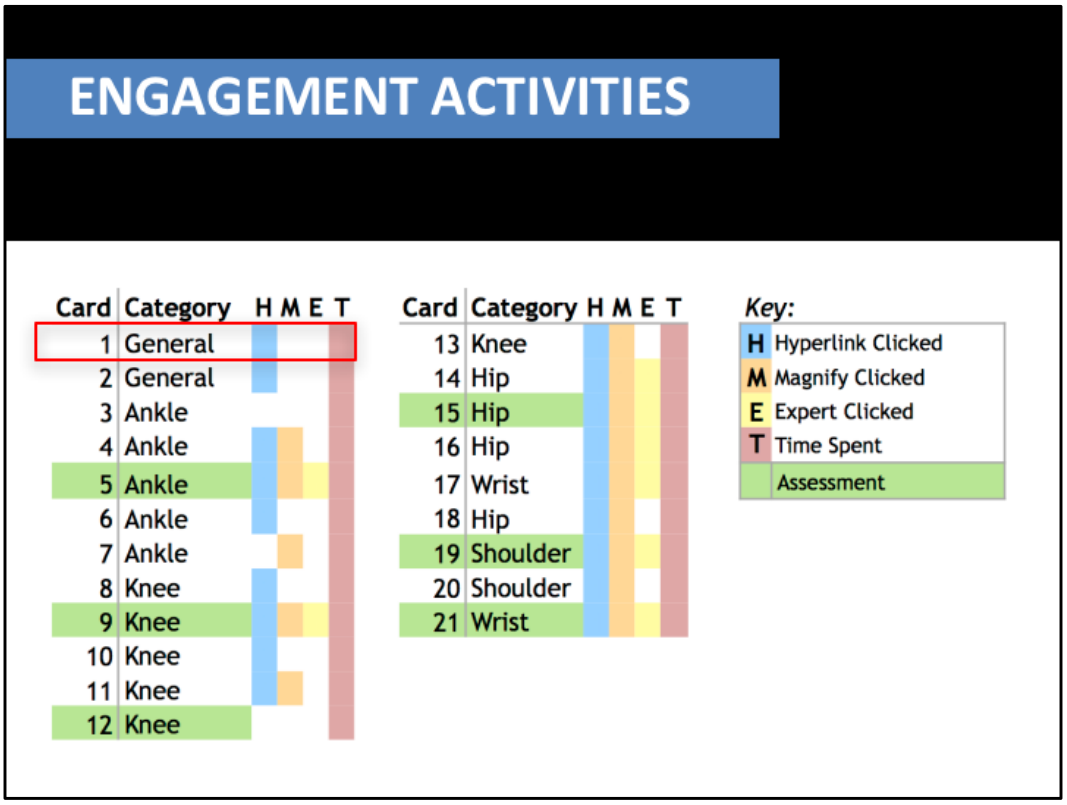


Here is a map of the content for the course we studied, with each card's number and topic category
Assessments are highlighted in green.

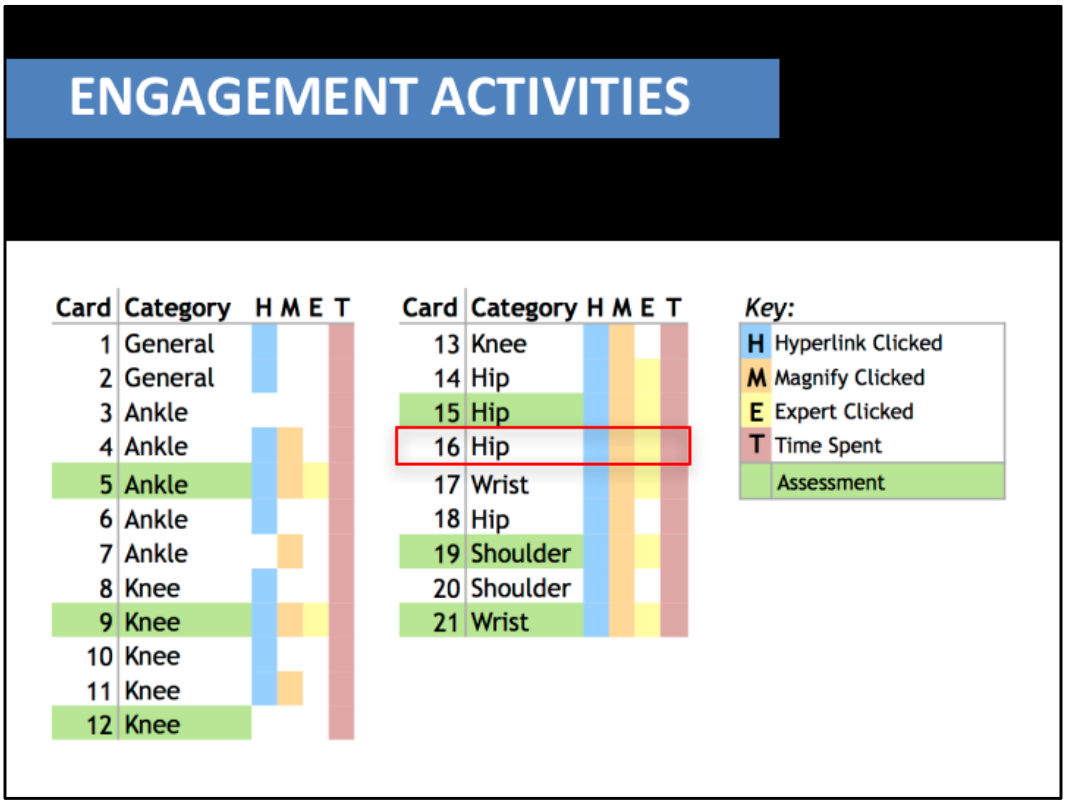
ENGAGEMENT ACTIVITIES



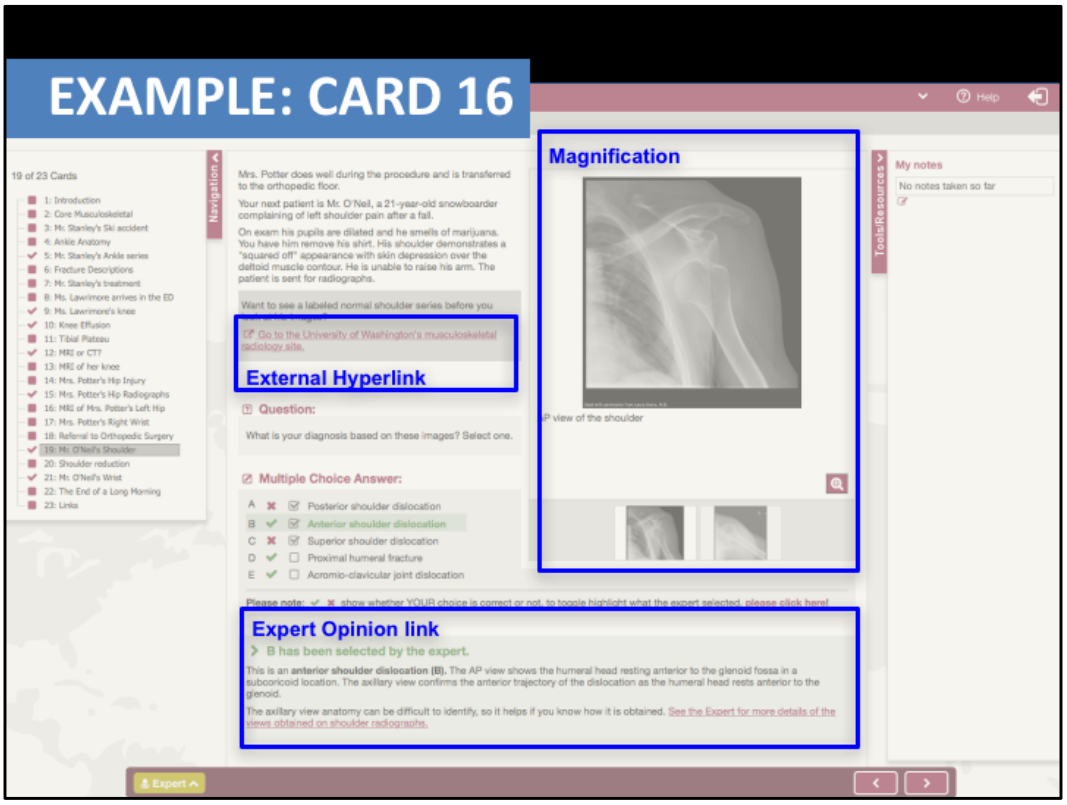
Here is where those engagement activities were distributed.
Not every engagement activity was available on every card



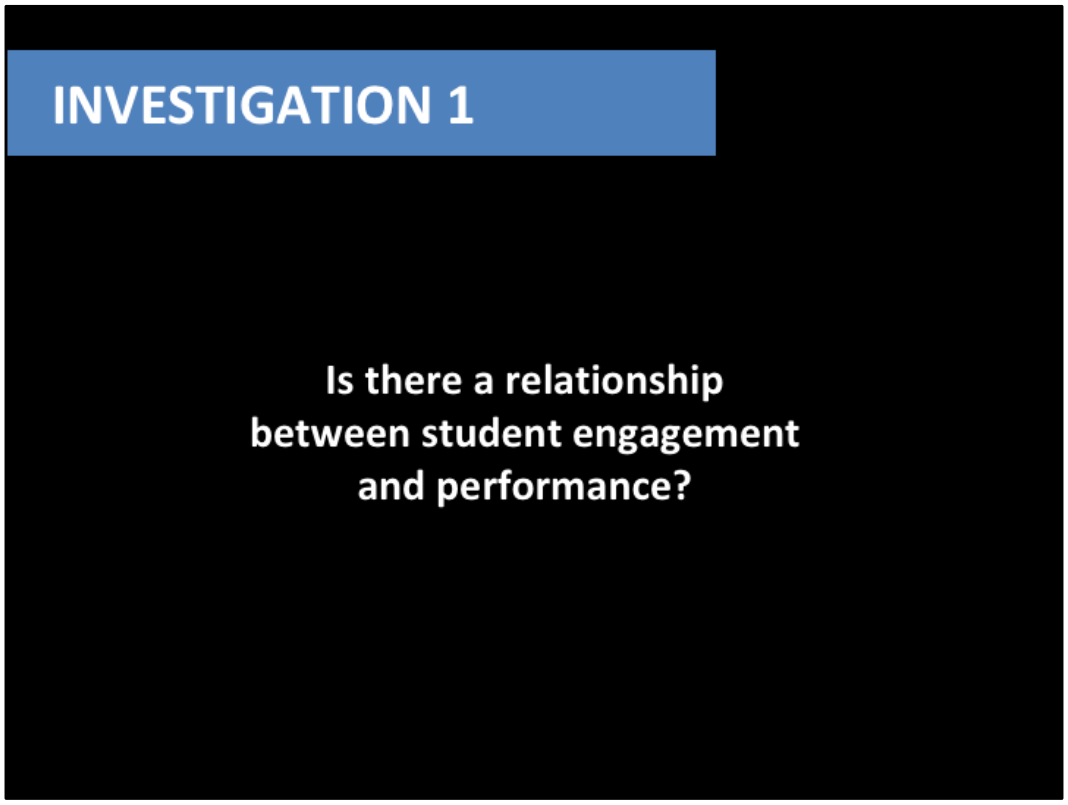
For example, card one only had an external hyperlink on it, so we only had insight into that click and time spent



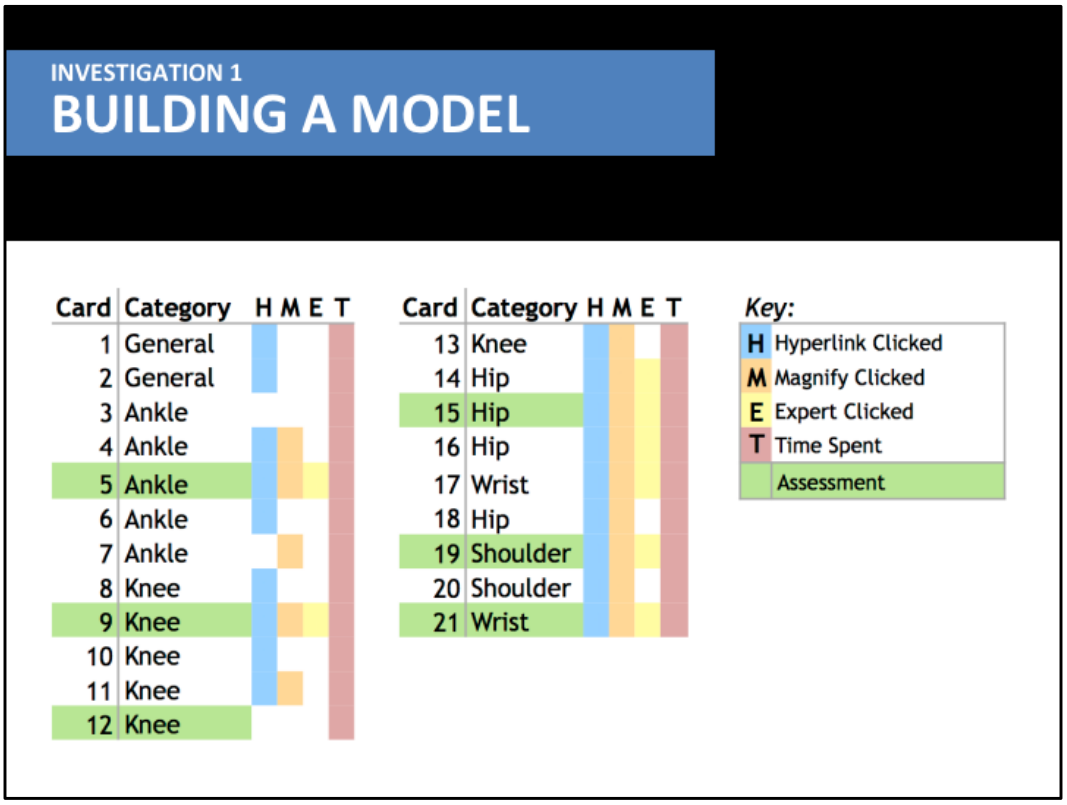
But card 16 had all three types of activity, plus time spent


EXAMPLE: CARD 16
Magnification
External Hyperlink
Expert Opinion link

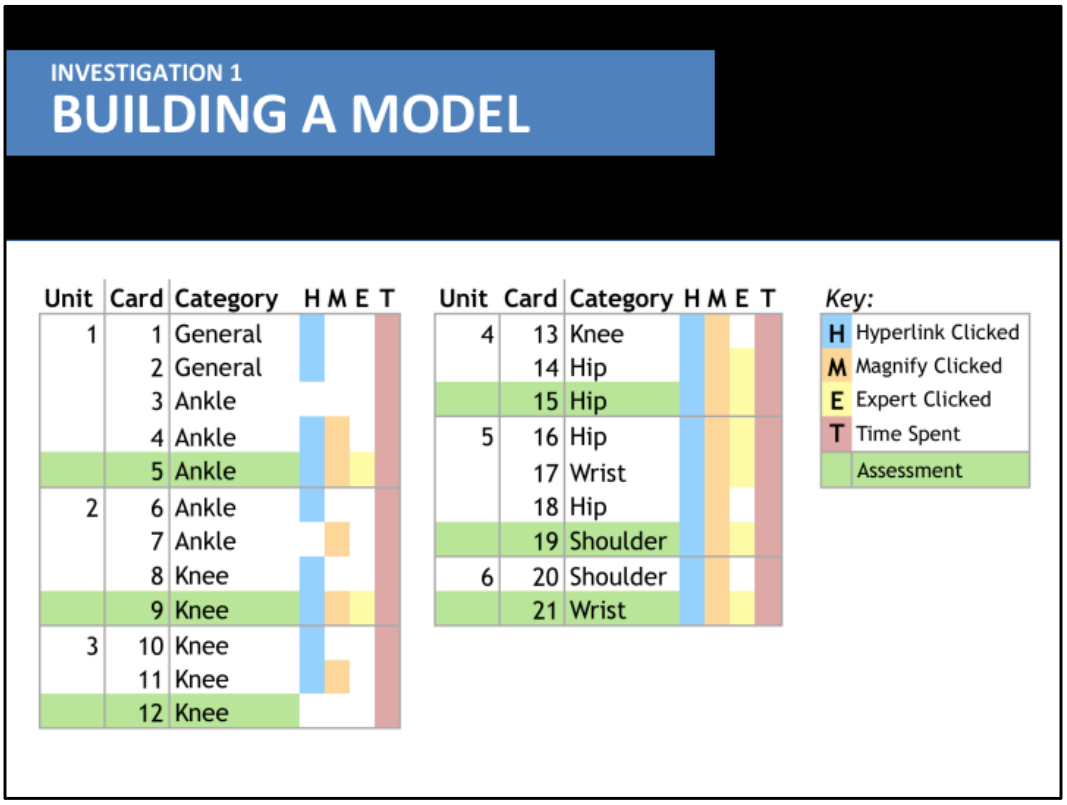
Shown here



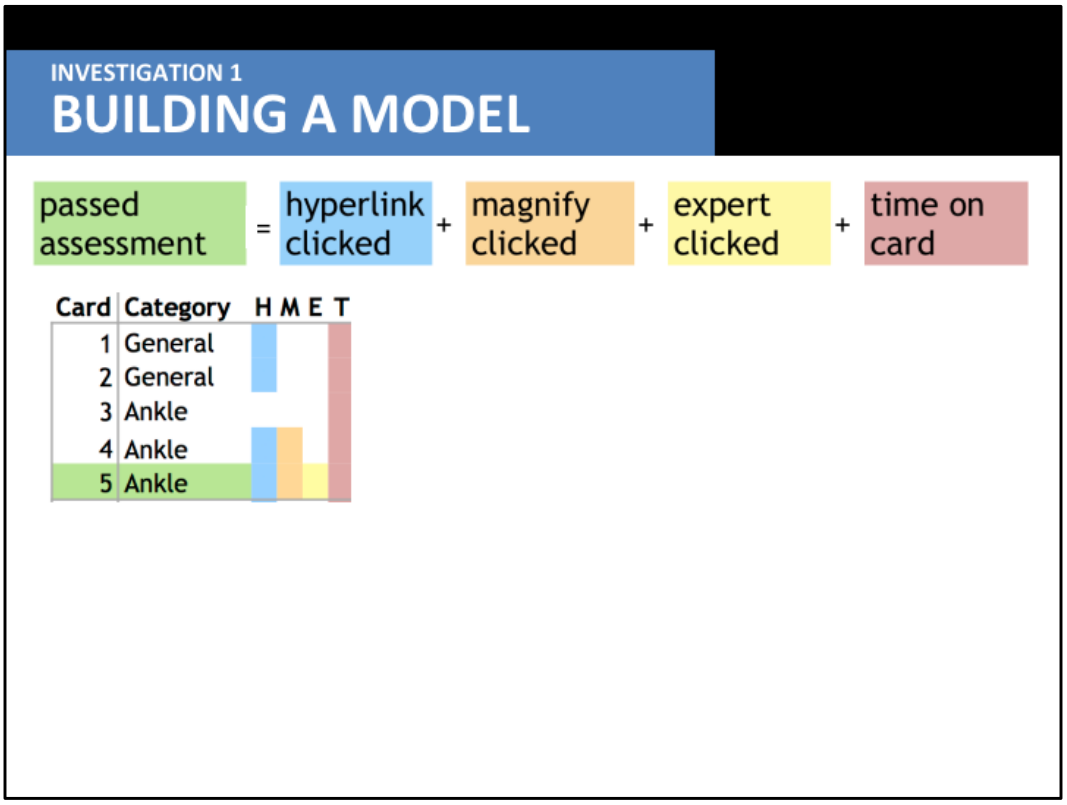
First investigation was to test our assumption that more engaged students performed better



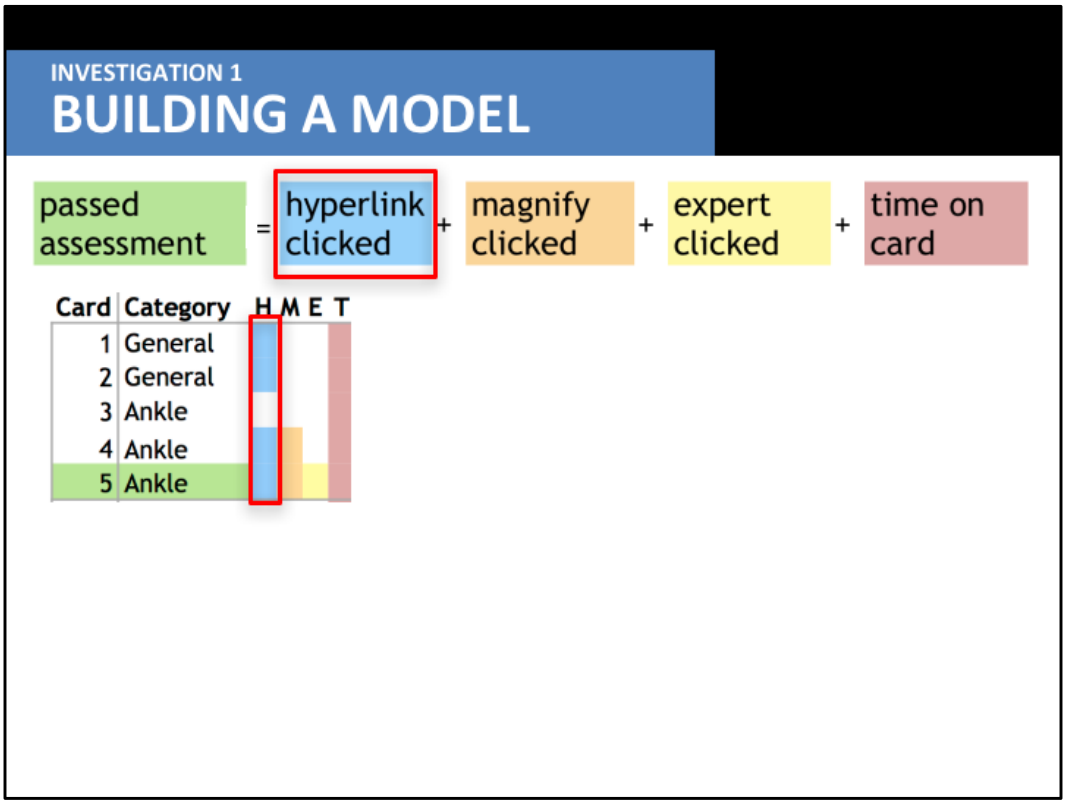
To answer that question, we broke up the course into units...



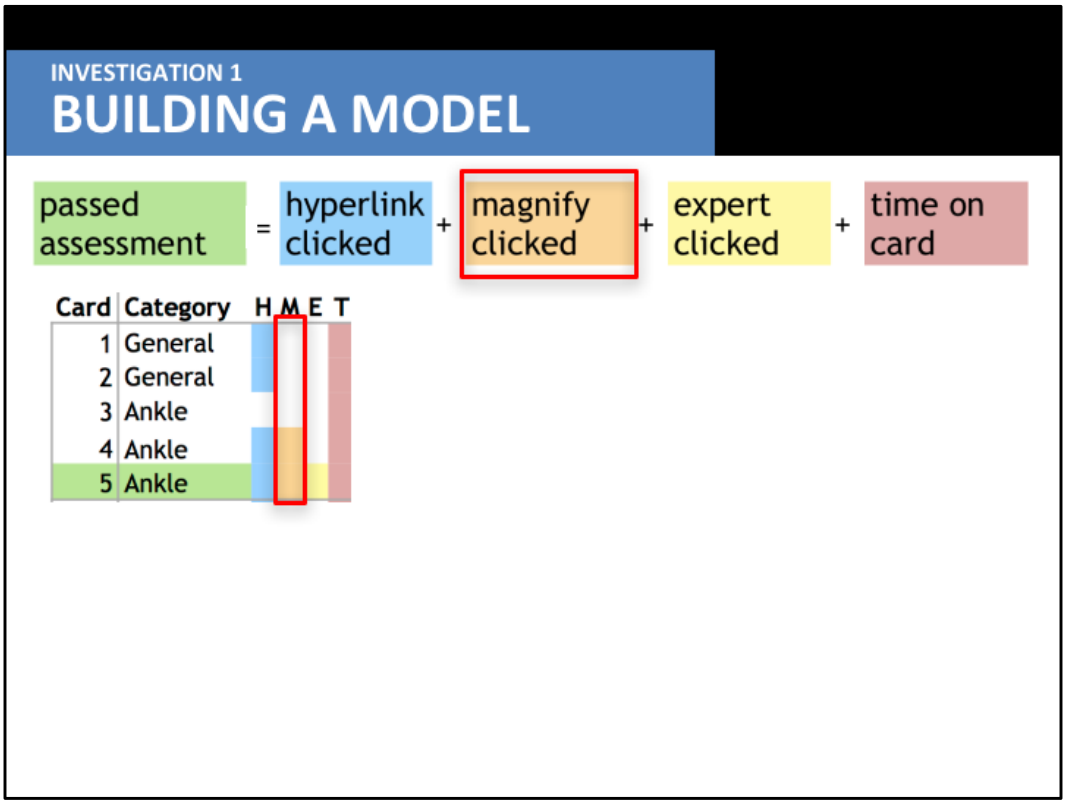
And for each unit...



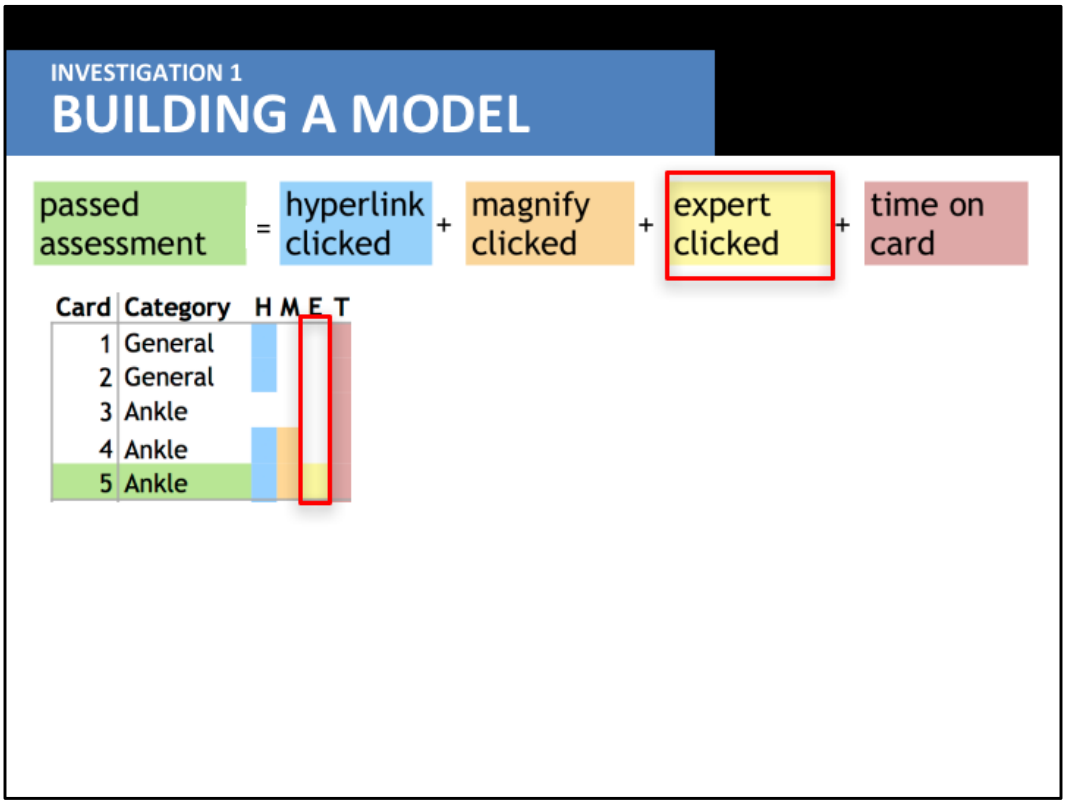
The model predicts whether a student will pass the end-of-unit assessment given...



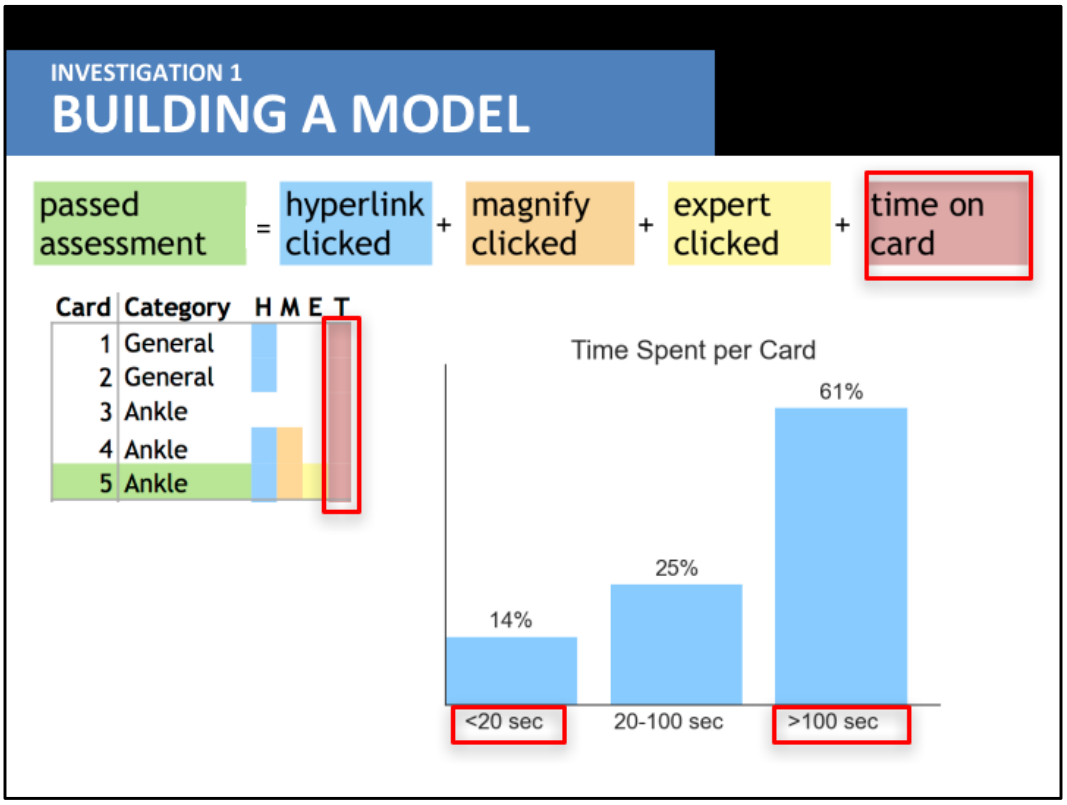
Given whether the student clicked on any of the available links,



Any of the magnify image buttons,

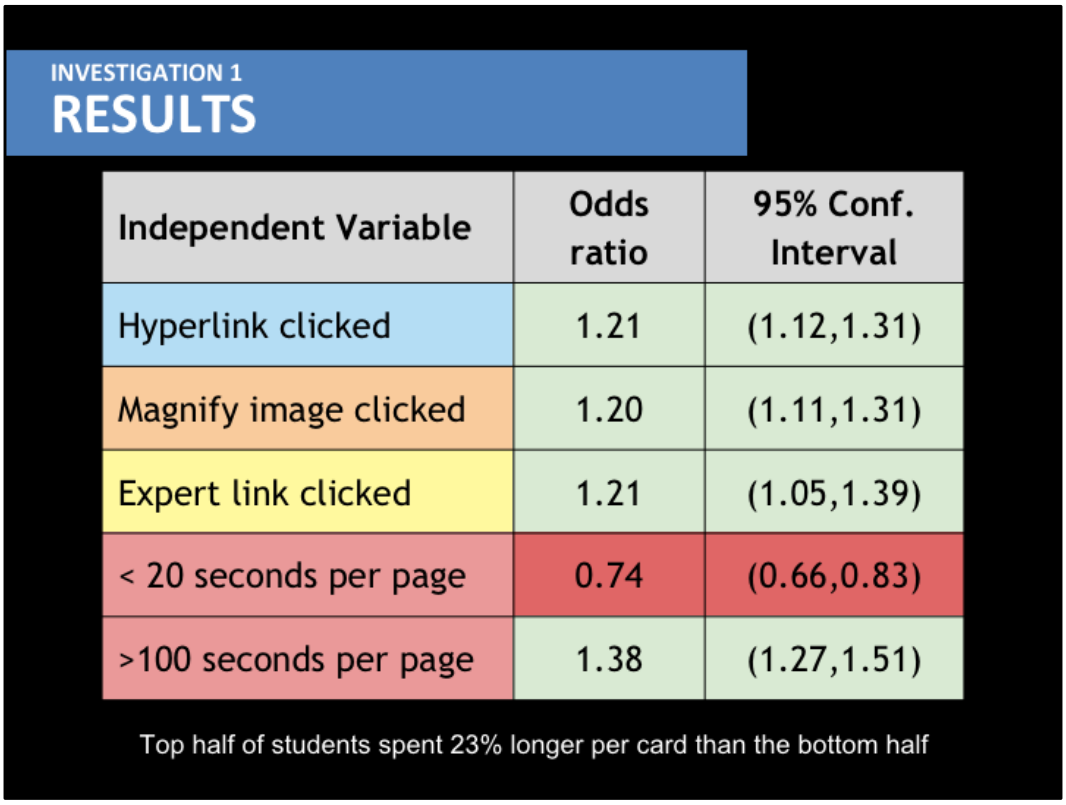


Any of the expert links



And how much time the student spent on each card.

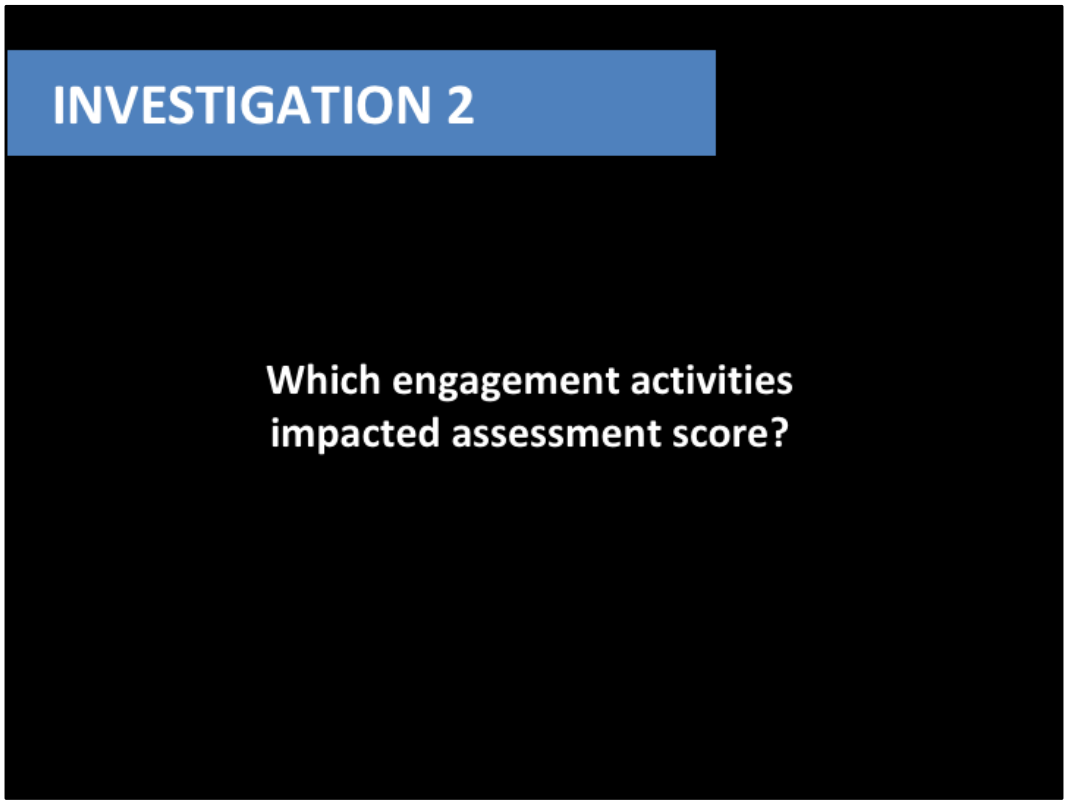
But since we expected a nonlinear relationship between time and performance,
we split time spent on card into bins:
and had indicators for each



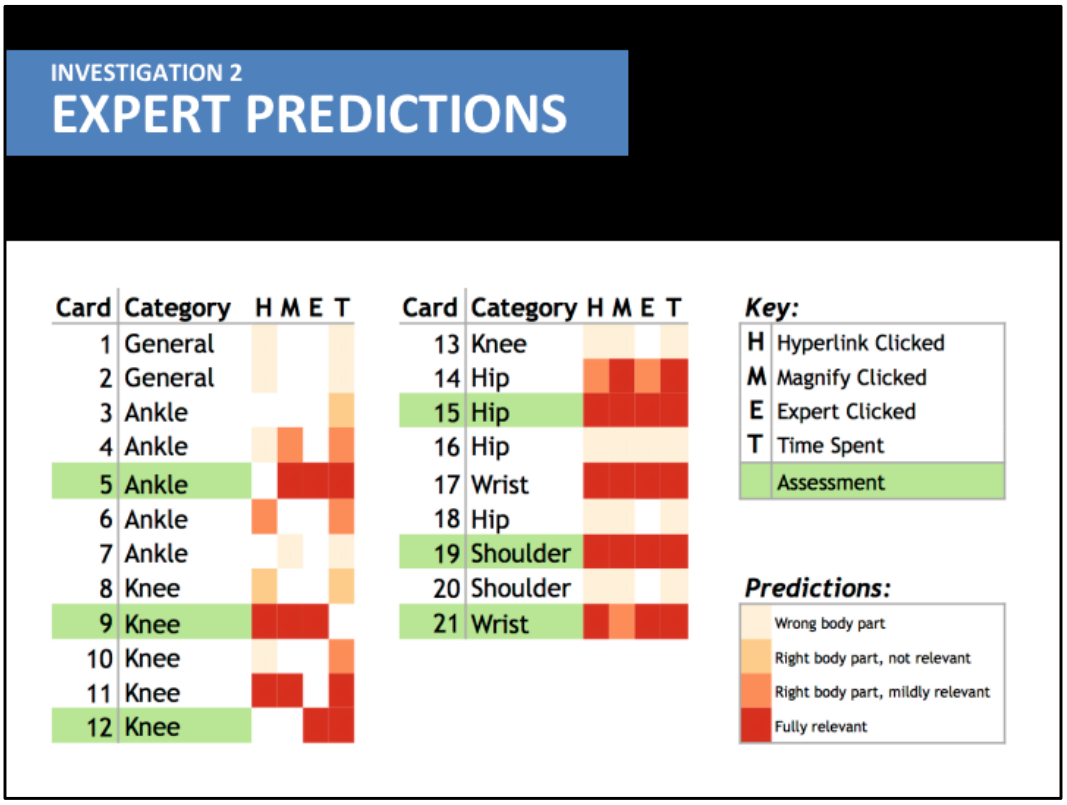
We tried several models, including decision trees and logistic regression for various transformations of the data,
but the best-performing model was logistic regression, [with AUC of 0.594].

***** All features statistically significant**
As expected, engagement and performance were related.
Students that rushed through the cards had lower performance on assessments.

But that's relatively obvious. What we really want to know is, which of the materials provided are useful to the students?

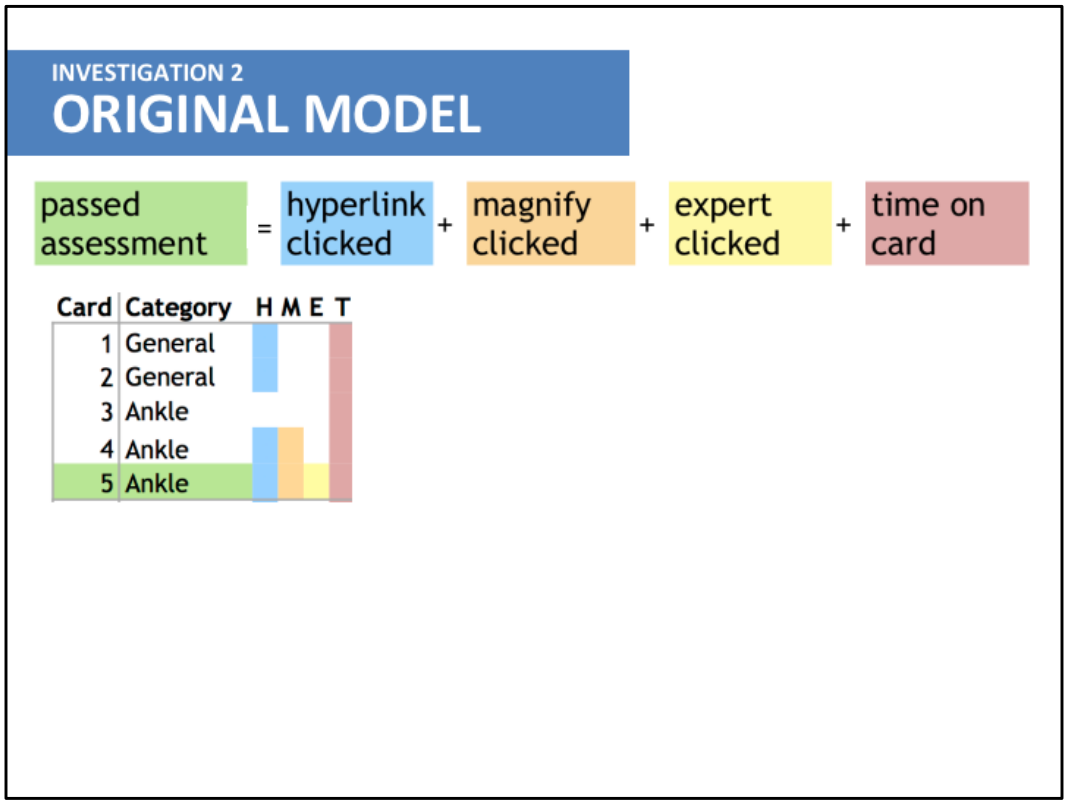


Yes studying helps pass tests, but which materials are useful and which are not? Which materials should be removed and replaced with others?

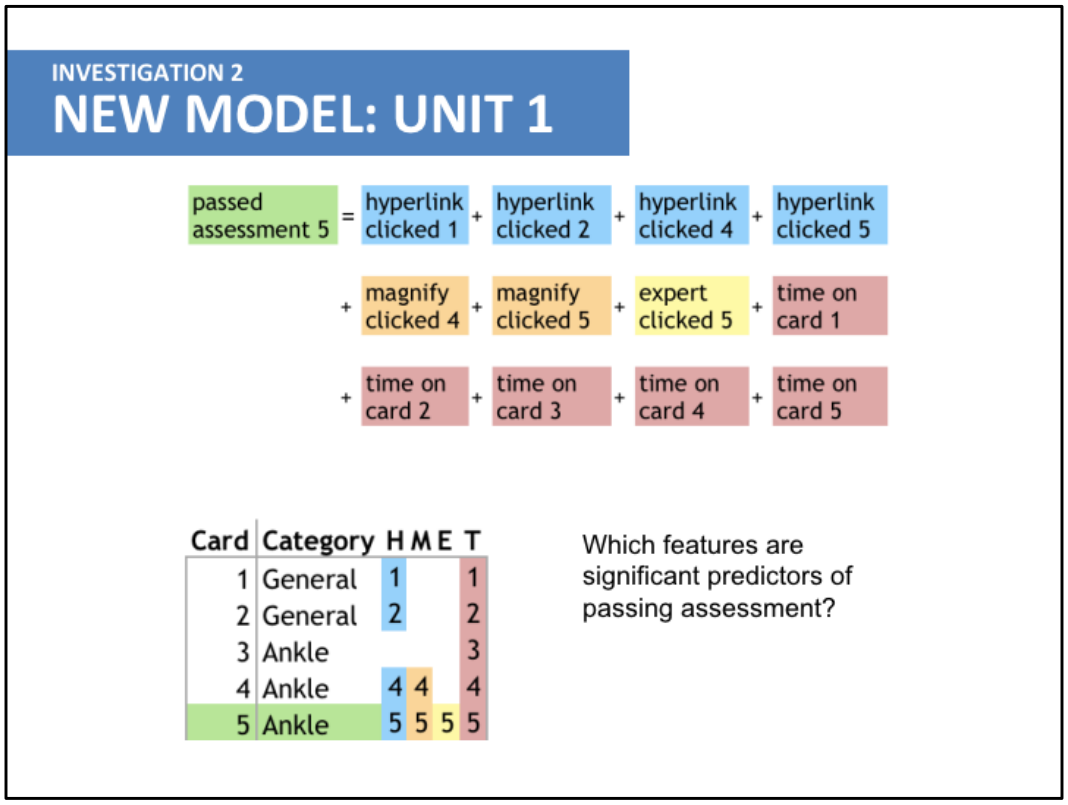


First Dr. Pusic provided his expert opinion,
predicting which materials would be most useful to students in answering subsequent assessment questions.

Darker colors are expected to be more relevant
For example he predicted that card 16's materials, which were about the Hip, would not be useful for the assessment on card 19, which is about the shoulder.



Like before, we broke up the course into units...



But this time built a separate model for each unit.

Unlike the first model, which lumped together any engagement activities,

Now we are looking for specific activities that contribute to performance

To do that, we consider each event separately so that we can see its impact

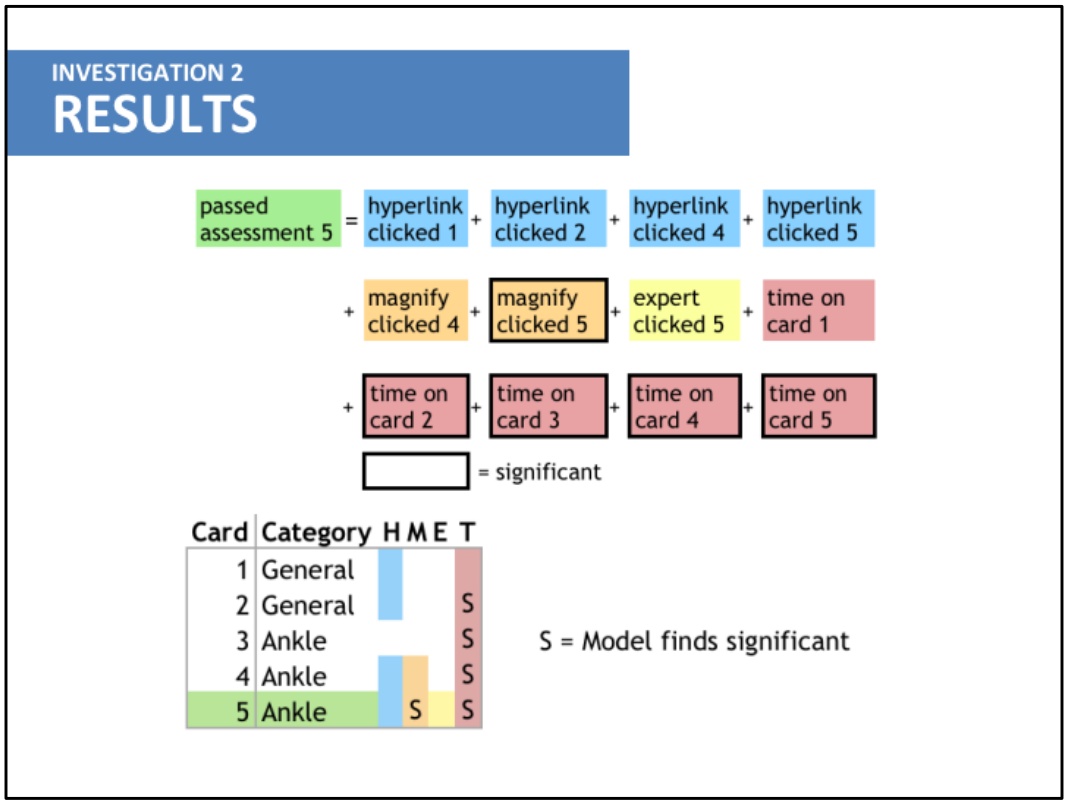
Investigation 2: Procedure

1. Run lasso-regularized logistic regression using all activities before assessment card

2. Find largest regularization parameter that is close to maximum cross-validation AUC

3. Re-run logistic with remaining variables

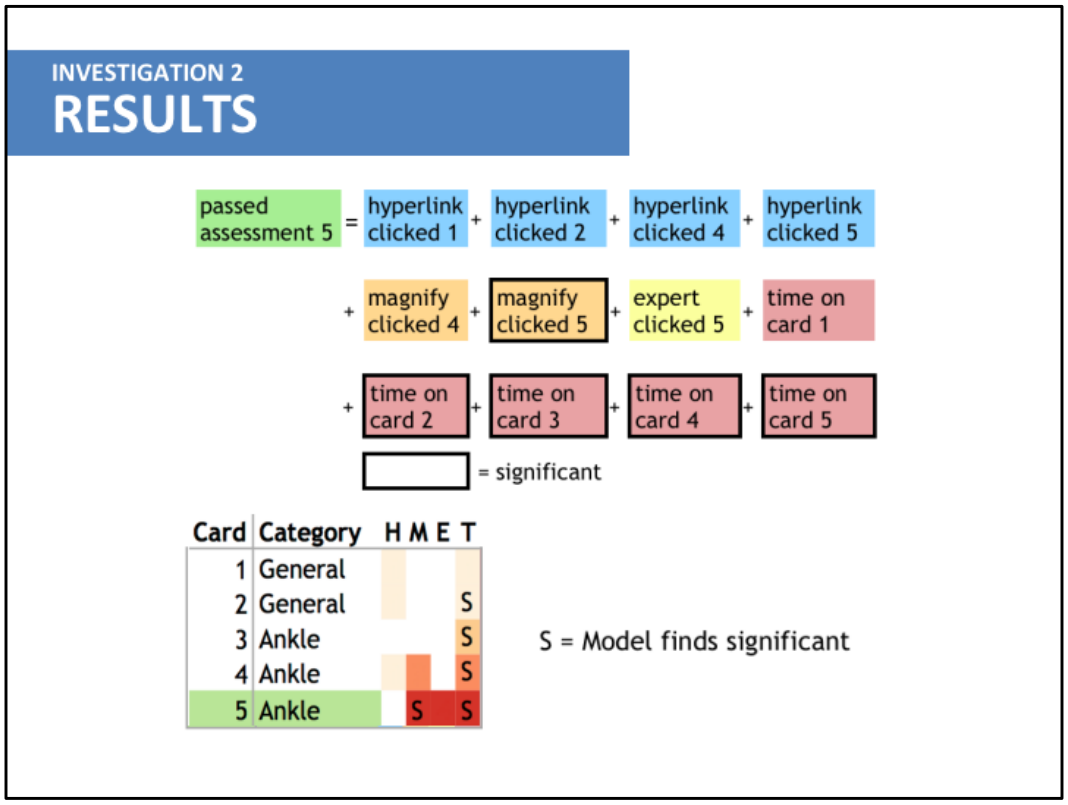
4. Return variables that have significant impact with p-value < 0.05



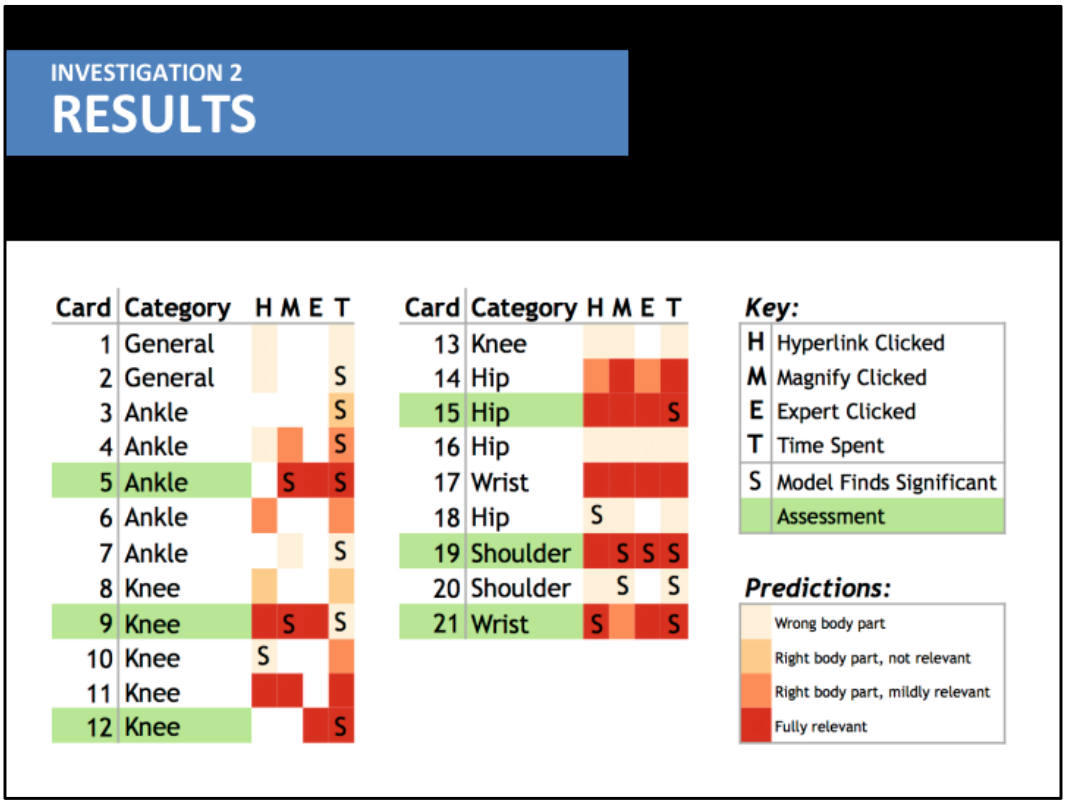
Then after running a model, we record which activities were significantly correlated with passing the assessment

In this unit's model,
we find that students who clicked the magnifier on card 5,
or spent more time on cards 2,3,4,5,
were more likely to pass the assessment on card 5.

But the other engagement activities were not significant predictors of passing probability



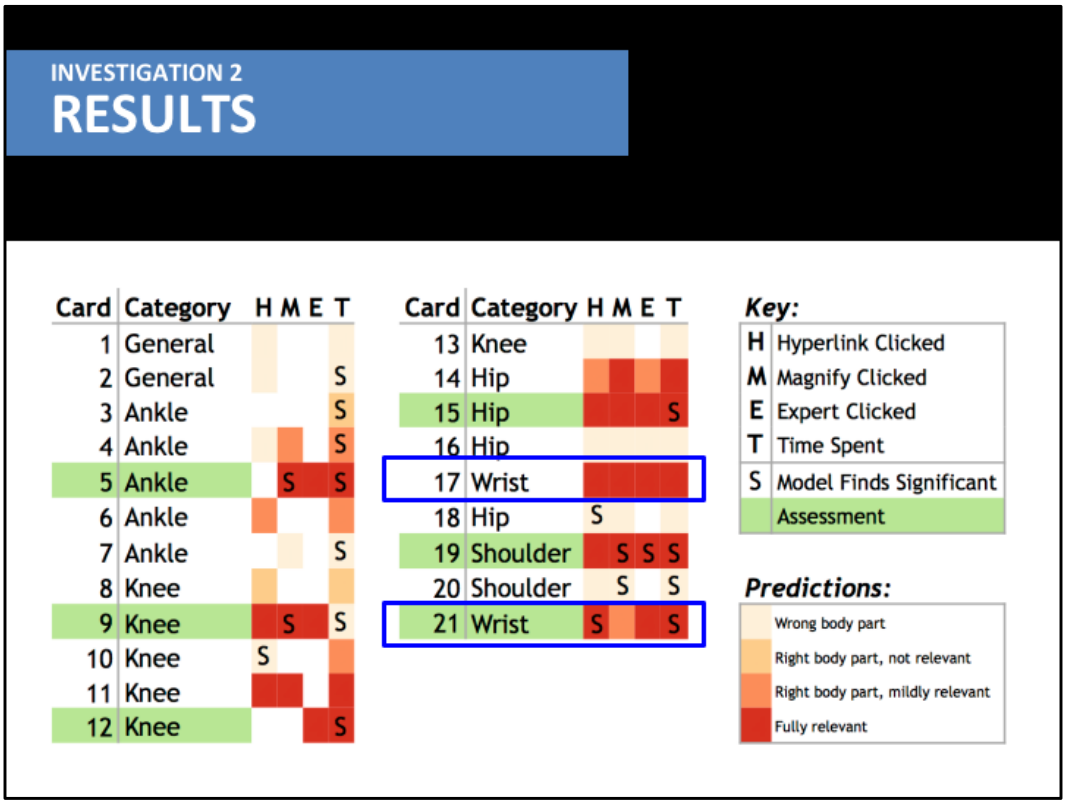
Comparing the model's results to Dr. Pusic's predictions shows where the predictions were and were not supported by the data.



Repeating that step for all units, we have this chart.

We can use these insights about what materials are NOT predictive of good performance,
and can recommend that instructional designers replicate what is working and replace what isn't

Notable Observations:...



19. Dr. Pusic's predictions for card 19 were consistent with model results.
You need to engage with that material in order to answer the question correctly

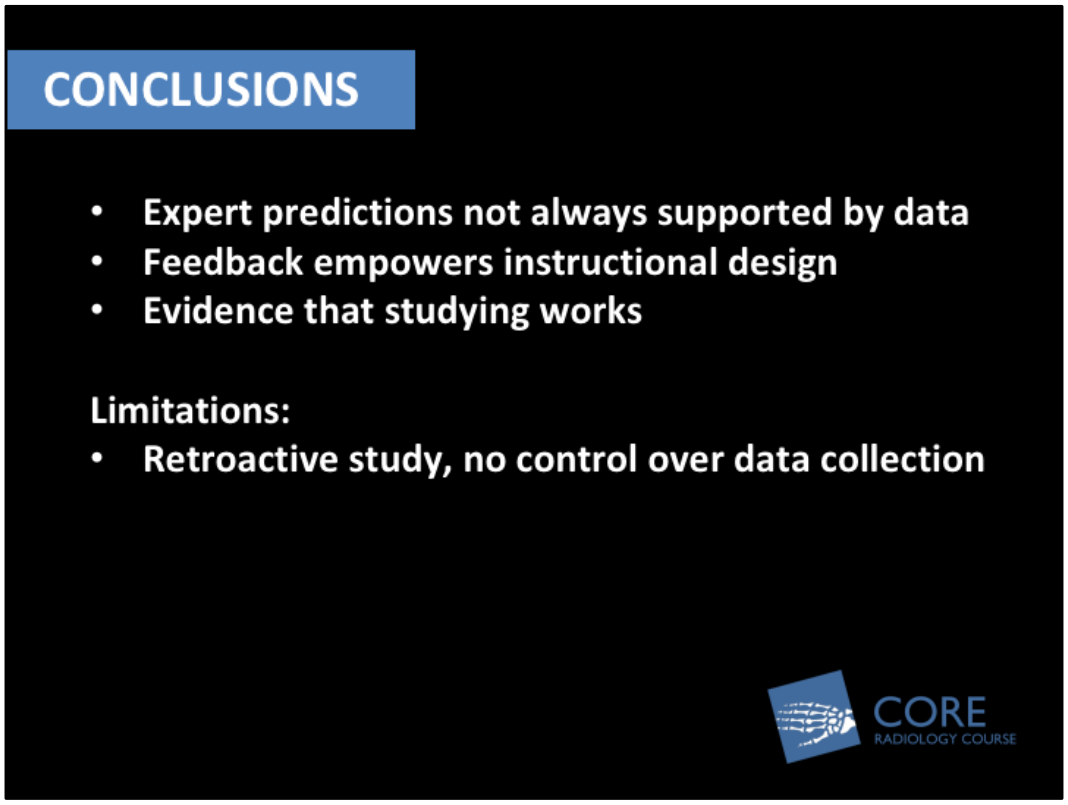
But
17 - Card 17, which is about the wrist, should be predictive of performance on card 21;
but perhaps there was too many cards between it and card 21 for students to see the relevance

1 - Card 5
Cards 1 and 2 comprise general content, and do not cover the assessment topic of ankles. These were understandably not predictive.

2 - Card 9
Engagement with card 6 and 8 was not associated with improved performance. Instructional designers should reconsider their inclusion.

3 - Card 12
Card 11 was expected to be fully relevant by the content expert, but the model did not consider its content predictive.

4 - Card 15
Few predictive variables were observed in this unit. Only time spent on the assessment card proved significant.



Conclusions/Strength of Innovation: Our intention was to demonstrate the merits of learning analytics within the online context, giving educators a new tool for improving experiences in educational online learning environments. Results of this analysis, where the data from thousands of learners are summarized, can serve as feedback to instructional designers as to which interaction elements are effective. It may also be useful to show students themselves evidence that there is a statistically significant relationship between engaging with the material and performing well on assessments.

DISCLOSURE

Grant/Research support and historical MedU data provided by:

http://www.med-u.org

THANK YOU

Special Thanks to MedU,
for research support and historical data.

For further information, feel free to contact us at:

Matt Cirigliano, MS matt.cirigliano@nyu.edu
NYU Steinhardt - CREATE Lab

Charlie Guthrie, BA charles.d.guthrie@gmail.com
NYU Center for Data Science

Martin Pusic, MD, PhD martin.pusic@nyumc.org
NYU School of Medicine
Division of Education Quality and Analytics (DEQA)