Synchronized eye movements predict test scores in online video education

Jens Madsen[a,1], Sara U. Júlio[a], Pawel J. Gucik[a], Richard Steinberg[b,c], and Lucas C. Parra[a]

[a]Department of Biomedical Engineering, City College of New York, New York, NY 10031; [b]School of Education, City College of New York, New York, NY 10031; and [c]Department of Physics, City College of New York, New York, NY 10031

Edited by David J. Heeger, New York University, New York, NY, and approved November 9, 2020 (received for review August 10, 2020)

Experienced teachers pay close attention to their students, adjusting their teaching when students seem lost. This dynamic interaction is missing in online education. We hypothesized that attentive students follow videos similarly with their eyes. Thus, attention to instructional videos could be assessed remotely by tracking eye movements. Here we show that intersubject correlation of eye movements during video presentation is substantially higher for attentive students and that synchronized eye movements are predictive of individual test scores on the material presented in the video. These findings replicate for videos in a variety of production styles, for incidental and intentional learning and for recall and comprehension questions alike. We reproduce the result using standard web cameras to capture eye movements in a classroom setting and with over 1,000 participants at home without the need to transmit user data. Our results suggest that online education could be made adaptive to a student's level of attention in real time.

online education | eye tracking | intersubject correlation

In a classroom the level of attention is quite variable (1, 2). An overt indicator of attention is the point of gaze (3, 4). When students are not following the relevant teaching material, there is a good chance that they are not paying attention and that they will perform poorly in subsequent exams. Experienced teachers know this and adjust the interaction with students accordingly (5). During online education this immediate feedback is lost. Here we suggest that standard web cameras could be used to monitor attention during online instruction based on the student's eye movements.

In the context of online media eye tracking has been used extensively to evaluate user interfaces, advertising, or educational material (6, 7). In education research eye tracking has been used to improve instructional design, determine the level of learner expertise, or purposefully guide eye movements during instruction (8). These studies often focus on the content of eye fixations in static media, to determine, for example, whether users look at a specific graphic or whether they read a relevant text (9, 10). This approach requires detailed analysis and interpretation of the specific content and cannot be used routinely to evaluate individual students. Evaluating the content of eye fixations is particularly complicated for dynamic stimuli such as instructional video, which is increasingly abundant online. Here we focus on dynamic video and whether students "follow" that dynamic content, in the literal sense of following with their eyes.

Previous studies have shown that eye movements are correlated across subjects during video presentation (11, 12). This intersubject correlation (ISC) of eye movements is elevated for dynamic, well-produced movies and video advertising (13–16) and is affected by the viewing task (17). A variety of eye-tracking measures have been used in educational research (18). However, the observation that eye movements are synchronized across subjects has not been widely explored in the context of education. In particular, it has not been established yet whether intersubject correlation of eye movement depends on attention or whether it is predictive of learning. Much of our eye movements during video seems to be driven by the visual dynamic (14, 19), resulting in similar scan paths even when movies are presented backward in time (11). Thus, a remarkable fraction of eye movements seems to be guided by "bottom-up" processing of salient visual events in the video (19, 20).

We hypothesize that typical online instructional videos synchronize eye movements across students; however, the level of synchrony depends on whether students are paying attention. Therefore, the correlation of eye movement between subjects should be predictive of retention of the material presented in the video. The alternative hypothesis is that the stimuli drive eye movements without engaging a student's mind meaningfully in the material. One may also argue that static stimuli, while not reliably guiding eye movements, may nonetheless engage students' minds (21, 22).

We test this hypothesis by measuring ISC of eye movements and pupil size, recorded while a diverse group of students watch short informal instructional videos typically found online. Consistent with the hypothesis, we find significant intersubject correlation, which drops in magnitude when viewers are distracted by a secondary task. Additionally, ISC of individual students is predictive of individual performance in a subsequent test of recall and comprehension. To determine the robustness of these findings, we repeat the experiment for different learning

Significance

Education is increasingly delivered online, but are students actually paying attention? Here we demonstrate that efficacy of video instruction can be assessed remotely with standard web cameras. Specifically, we show that attentive students have similar eye movements when watching instructional videos and that synchronization of eye movements is a good predictor of individual learning performance. Measuring synchronization of eye movements while preserving privacy, as we have shown here, has the potential to make online education adaptive to attentional state and advance mechanistic studies on the efficacy of different online education formats. Attention has become a commodity online. With the increasing abundance of video content, remote sensing of attention at scale may be relevant beyond education, including entertainment, advertising, and politics.

Author contributions: J.M. and L.C.P. designed research; J.M., S.U.J., P.J.G., R.S., and L.C.P. performed research; L.C.P. contributed new reagents/analytic tools; J.M. analyzed data; and J.M. and L.C.P. wrote the paper.

Competing interest statement: L.C.P. and J.M. are listed as inventors in a related pending patent application. L.C.P. was affiliated with Neuromatters Corp., KCore Analytics Inc., and 3Finches Inc. at the time of the study.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jmadsen@ccny.cuny.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2016980118/-/DCSupplemental.

Published January 28, 2021.

scenarios and instructional videos produced in different styles. Finally, we replicate the results with remote students, using subjects' own computers to capture their eye movements, without the need to transfer data from the user, thus preserving online privacy.

Results

Effects of Attention on Eye Movements during Video Presentation. To test the hypothesis that synchronization of eye movements depends on attentional state, we recruited a diverse group of subjects (n = 88) to participate in a series of experiments where they were asked to watch five or six short instructional videos in the laboratory while we monitored their eye movements. The videos covered a variety of topics related to physics, biology, and computer science (*SI Appendix*, Table S2). The videos reflected the most common contemporary formats, which feature a teacher writing on a board, or more modern storytelling using animations or the popular writing-hand style. A first cohort of subjects (n = 27, 17 females, age 18 to 53 y, mean [M] = 26.74, standard deviation [SD] = 8.98) watched five short instructional videos, and after each video they took a test with questions related to the material presented in the videos. In the first cohort, subjects were told to expect this subsequent test. After watching the videos and answering questions they watched the videos again. To test for attentional modulation of ISC, in the second viewing subjects performed a serial subtraction task (count silently in their mind backward in steps of seven starting from a random prime number between 800 and 1,000). This is a common distraction task in visual attention experiments (23). During the first attentive viewing eye movements of most subjects are well correlated (Fig. 1*A*). As predicted, during the second one, distracted viewing eye movements often diverge (Fig. 1*B*). The same appears to be true for the fluctuations of pupil size. To quantify this, we measure the Pearson's correlation of these time courses between subjects. For each student we obtain an ISC value as the average correlation of that subject with all other subjects in the group. We further average over the three measures taken, namely, vertical and horizontal gaze position as well as pupil size. This ISC is substantial during the normal viewing condition (Fig. 1*C*; ISC median = 0.35, interquartile range [IQR] = 0.12, across videos) and decreases in the second distracted viewing (ISC median = 0.12, IQR = 0.18). Specifically, a three-way repeated-measures ANOVA shows a very strong fixed effect of the attention condition ($F(1, 231)$ = 749.06, P = $1.93{\cdot}10^{-74}$), a fixed effect of video ($F(4, 231)$ = 32.29, P = $2.23{\cdot}10^{-21}$), and a random effect of subject ($F(26, 231)$ = 9.21, P = $1.62{\cdot}10^{-23}$). For a replication of these results on two other experiments see *SI Appendix*, section S1. This confirms the evident variability across films and subjects. The predicted effect of attention, however, is so strong that despite the variability between subjects one can still determine the attention condition from the ISC of individual subjects (Fig. 1*C*). Specifically, by thresholding on ISC one can determine with high accuracy the attentional state of the subject (area under the receiver–operator curve of $Az = 0.944 \pm 0.033$ – mean ± SD over videos). To determine which time scale of the eye dynamic dominates this intersubject correlation we computed ISC resolved by frequency (Fig. 1*D*). We find that ISC and its modulation with attention are dominant in a time scale of 0.1 to 33 s for eye movements (0.03 to 8.8 Hz) and 0.2 to 33 s for pupil size (0.03 to 5 Hz).

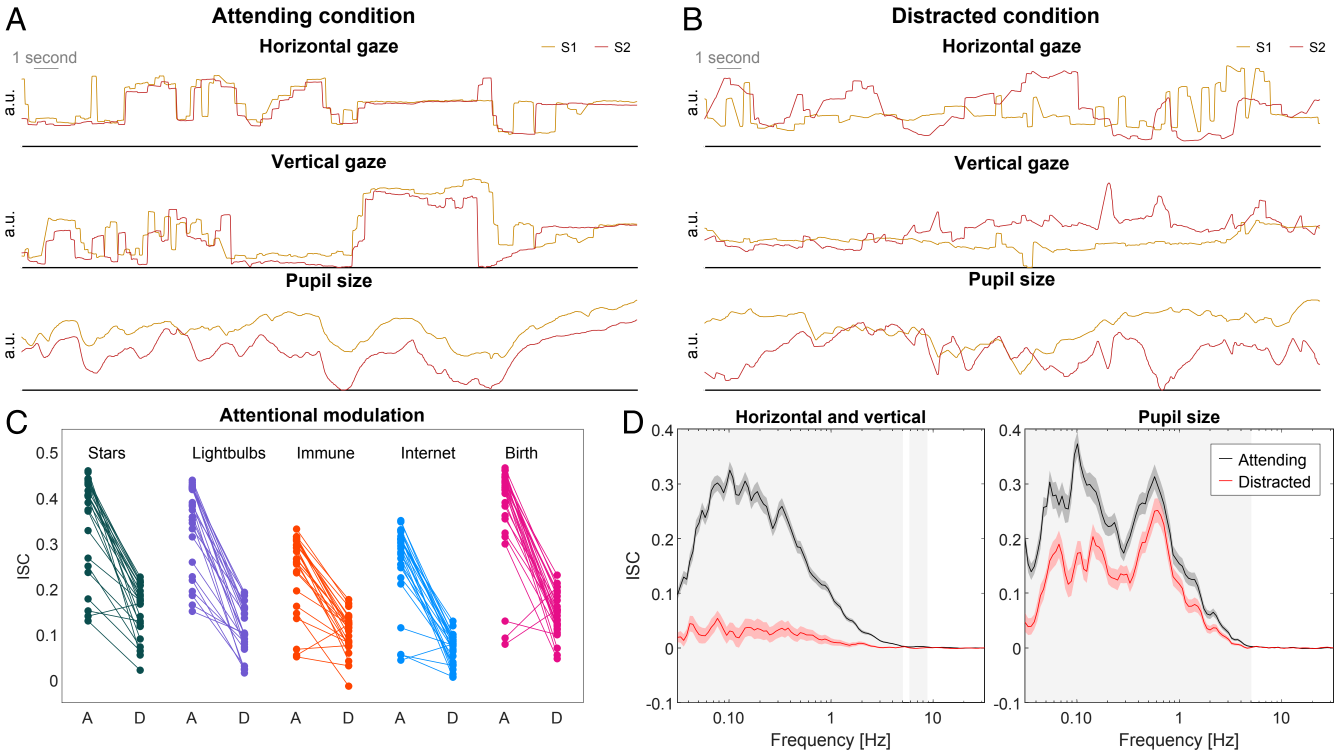

Fig. 1. Intersubject correlation of eye movements modulated by attention when watching instructional videos. (*A*) Two subjects' gaze position and pupil size follow each other during attentive viewing. (*B*) The same two subjects viewing the same segment of video while distracted by a counting task. (*C*) For each subject, ISC is measured as the mean correlation of vertical and horizontal gaze position and pupil size with that of other subjects. Values for each subject are shown as dots for all videos in experiment 1. Each dot is connected with a line between two different conditions, namely, when subjects were either attending (A) or distracted (D) while watching the video. (*D*) ISC for the attentive and distracted conditions resolved by frequency, i.e., computed on band-pass filtered eye movements and pupil size. Each ISC value is averaged over the five videos and all subjects. In *D*, *Left* and *Right* significant differences between attending and distracted conditions are established using a cluster permutation test (gray shaded area indicates $P < 0.01$).

Correlated Eye Movements and Pupil Size as Predictors of Test Scores. In the previous experiments we confirmed the hypothesis that if subjects are distracted the ISC of eye movements and pupil size is reduced. Given the well-established link between attention and memory we therefore expect that ISC will be predictive of how much each subject retained from the instructional video. We tested this hypothesis by quizzing subjects after they had watched the video using a short, four-alternative forced-choice questionnaire (11 to 12 questions per video). Students that watched the video performed significantly better than naïve students (65.2 ± 18.8% versus naïve: 45 ± 14.6%; $t(56) = 5.37$, $P = 1.58 \cdot 10^{-6}$; see *Materials and Methods* for details). Importantly, we find a strong correlation between ISC and test scores across subjects for all videos we tested (Fig. 1*B*; $r = 0.61 \pm 0.06$, SD across five videos, $P < 3.60 \cdot 10^{-3}$). This is true for ISC of eye movement and pupil size alike, even when luminance fluctuations are regressed out from the pupillary response (*SI Appendix*, Fig. S6). Evidently, subjects with lower ISC performed poorer on the tests (e.g., subject 3 in Fig. 2*A*). Inversely, subjects with more correlated eye movements obtain higher test scores (e.g., subjects 1 and 2 in Fig. 2*A*). This suggests that subjects who did not follow the visual dynamics of the video with their eyes were not paying attention and as a result their test scores were lower (see *SI Appendix*, section S2 for replication of these results on two other laboratory experiments). This causal interpretation is consistent with a statistical model of the data (*SI Appendix*, section S4). However, the present study is only observational and the source of the correlation observed here between ISC and test performance remains undetermined.

A classic approach to predicting recall of a visual stimulus is to measure overt attention toward the stimulus, for instance, by measuring the number of fixations or their duration (23, 24). Indeed, in the present data the fraction of time subjects had their eye on the video also correlated with subsequent test scores, but the effect was not as strong as for the ISC of eye movements (*SI Appendix*, section S9 and Fig. S7).

Different Learning Scenarios. To test for the robustness of the effect we repeated the experiment, but this time subjects did not know that they would be quizzed on the content of the videos (Fig. 2*C*). The two scenarios thus constitute intentional learning and incidental learning which are known to elicit different levels of motivation (25). As expected, we find a higher ISC in the intentional learning condition (ISC median = 0.32, IQR = 0.12, n = 27) compared to the incidental learning condition (ISC median = 0.317, IQR = 0.06, n = 30) (two-tailed Wilcoxon rank sum test: $z = 2.82$, $P = 4.78 \cdot 10^{-3}$). This suggests that lower motivation in the incidental learning condition resulted in lower attentional levels and thus somewhat less correlated eye movements and pupil size. In the intentional learning condition test scores were higher compared to the incidental learning condition (intentional learning score =

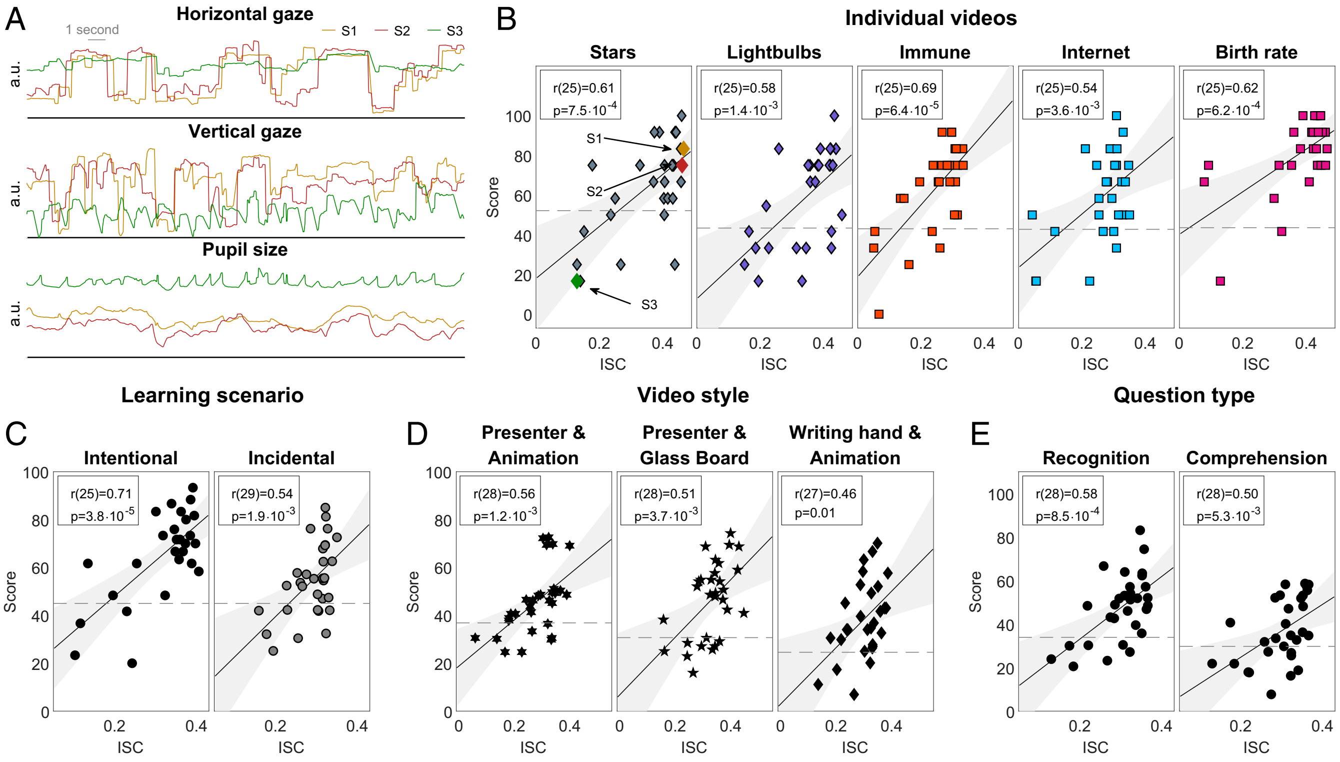

Fig. 2. Intersubject correlation of eye movements during instructional videos predicts learning performance. (*A*) Eye movements of three representative subjects as they watch "Why are Stars Star-Shaped?". Two high-performing subjects have similar eye movements and pupil size dynamic (S1 and S2). A third, low-performing student does not match their gaze position or pupil size (S3). (*B*) ISC of eye movements and performance on test taking (score) for each of five videos in experiment 1. Each symbol is a subject. The high- and low-performing subjects (subjects 1 to 3) from *A* are highlighted in color and with arrows for the stars video. Dashed lines represent performance of subjects naïve to the video. (*C*) Same as *B* but averaging over the five videos. The data were collected in two different conditions: during intentional learning (experiment 1) where subjects knew they would be quizzed on the material and during incidental learning (experiment 2) where subjects did not know that quizzes would follow the viewing. (*D*) Videos in three different production styles (experiment 3) show similar correlation values between test scores and ISC. Each point is a subject where values are averaged over two videos presented in each of the three styles. (See *SI Appendix*, Fig. S2 for results on all six videos.) (*E*) A similar effect is observed for different question types. Here each point is a subject with test scores averaged over all questions about factual information (recognition) versus questions requiring comprehension. ISCs were averaged over all six videos in experiment 3.

65.22 ± 18.75 points, $n = 27$; incidental learning score = 54.53 ± 15.31 points, $n = 31$; two-sample t test, $t(56) = 2.39$, $P = 0.02$, $d = 0.63$). This may reflect increased motivation or more simply the increased difficulty of having to answer all questions together after a longer time interval. Importantly, and again consistent with our hypothesis, in both cohorts there is a robust correlation between ISC and test scores (Fig. 2*C*; intentional, $r(25) = 0.61$, $P = 7.51 \cdot 10^{-4}$; incidental, $r(29) = 0.58$, $P = 5.87 \cdot 10^{-4}$).

Different Styles of Instructional Videos. We found a positive correlation between ISC and test scores for all five videos tested. The style of these five videos either consisted of animation (lightbulbs, immune, internet) or showed a hand, drawing figures (stars, birth). We wanted to test whether this effect is robust to other popular styles of informal instructional videos found on popular YouTube channels. To this end we performed an additional experiment on a new cohort of 30 subjects (experiment 3; 22 females, 8 males, age 18 to 50 y, mean = 25.50, SD = 8.05 y) where we selected six different videos in three different styles (two videos per style): a real-life presenter along with animation, a presenter writing on a glass board, and a writing hand with animation (see links to videos in *Materials and Methods*). Despite the different visual appearance and dynamic, we still find a strong correlation between ISC and test scores for all three styles (Fig. 2*D*: presenter and animation, $r(28) = 0.56$, $P = 1.2 \cdot 10^{-3}$); writing hand and animation, $r(27) = 0.46$, $P = 0.01$; presenter and glass board, $r(28) = 0.51$, $P = 3.7 \cdot 10^{-3}$)).

Recognition and Comprehension Questions. It is possible that attention favors recognition of factual information, but that questions probing for comprehension of the material would require the student to disengage from the video to process the content "offline." We therefore included in experiment 3 comprehension questions (32 of a total of 72 questions across the six videos; see questions in *SI Appendix*, Table S2). Overall subjects did similarly on the comprehension questions compared to the recognition questions (Fig. 2*E*) and we find a significant correlation with ISC for these comprehension questions ($r(28) = 0.50$, $P = 5.3 \cdot 10^{-3}$), and we again find a correlation with recognition performance ($r(28) = 0.58$, $P = 8.5 \cdot 10^{-4}$). These correlation values do not differ significantly (asymptotic z test after Fisher r-to-z conversion, $P = 0.68$), suggesting that comprehension and recognition are both affected by attention. Indeed, test scores for comprehension and recognition questions are significantly correlated across subjects ($r(28) = 0.52$ ($P = 3.02 \cdot 10^{-3}$)). Therefore, the hypothesized link between ISC and performance seems to be fairly robust, applying to different learning scenarios and various styles of educational video found online, as well as recognition and comprehension questions alike.

Capturing Eye Movements Online at Scale Using Standard Web Cameras. Thus far all experiments were performed in a laboratory setting with a research-grade eye tracker. To test the approach in a realistic setting we developed an online platform that can operate with standard web cameras. These cameras typically operate at lower sampling frequencies (<60 Hz), which should suffice as we showed that the relevant fluctuations of eye movements are below 10 Hz (Fig. 1*D*). The online platform relies on existing eye-tracking software that can run on any web browser (26), allowing us to reach a large scale of users. The software operates on the remote computer of the users and captures gaze position, but not pupil size as web cameras typically do not have the necessary spatial resolution. In one experiment we recruited 82 students (female = 21, age 18 to 40 y, mean = 19.6, SD = 2.7 y) from a college physics class to participate after their laboratory sessions using the desktop computers available in the classroom (experiment 4: classroom). In another experiment we recruited 1,012 participants (female = 443, age 18 to 64 y, mean = 28.1, SD = 8.4 y) on MTurk and Prolific. These are online platforms that assign tasks to anonymous subjects and compensate them for their work (experiment 5: at-home). The subjects used the webcam on their own computers, emulating the at-home setting typical for online learning. The gaze position data collected with web cameras are significantly noisier than using the professional eye tracker in the laboratory (Fig. 3*A*; see raw data in *SI Appendix*, section S6). To quantify this, we compute the accuracy of gaze position when subjects are asked to look at a dot on the screen (Fig. 3*B*). As expected, we find a significant difference in gaze position accuracy between the laboratory and the classroom (two-sample t test, $t(69) = -7.73$, $P = 6.3 \cdot 10^{-11}$) and a significant difference between the classroom and the at-home setting ($t(242) = -2.46$, $P = 0.01$). Despite this signal degradation we find a high correlation between the median gaze position data (across subjects) for laboratory and classroom data (horizontal gaze, $r = 0.87 \pm 0.04$; vertical gaze, $r = 0.75 \pm 0.04$) and laboratory and at-home data (horizontal gaze, $r = 0.91 \pm 0.04$; vertical gaze, $r = 0.83 \pm 0.04$).

Predicting Test Scores in a Classroom and at Home Using Web Cameras. To preserve online privacy of the users we propose to evaluate eye movements remotely by correlating each subject's eye movements with the median gaze positions (Fig. 3*A*). Instead of ISC with all members of the group, we thus compute the correlation with the median position locally, without the need to transmit individual eye position data (*Materials and Methods*). To compensate for the loss of the pupil signal we now also measure the correlation of eye movement velocity, which has been used previously to capture synchronous eye movements (16) (*Materials and Methods*). We combine these eye movement metrics by taking a weighted average of the vertical, horizontal, and velocity ISC (wISC) [following previous work on combining multiple ISC measures (27, 28); *Materials and Methods*]. We find that this wISC of eye movement robustly correlates with subsequent test scores (Fig. 3 and *SI Appendix*, Table S1) despite the lower quality of the gaze position data. In fact, the correlation of wISC with test scores for the classroom (Fig. 3*C*; $r = 0.46 \pm 0.16$, $P < 0.01$) is comparable to the values in the laboratory experiments ($r = 0.59 \pm 0.08$, all $P < 0.01$; compare to Fig. 2*B*). The at-home experiment had also highly significant correlation between wISC and subsequent test scores (Fig. 3*D*; $r = 0.47 \pm 0.08$, $P < 3.9 \cdot 10^{-8}$). The prediction error of the test score based on wISC is 14.6 ± 16.9% (median across videos, IQR across all videos and subjects), which is equivalent to 1.75 of 12 questions, and outperforms a naïve predictor based on mean performance (*SI Appendix*, Table S1). We can essentially predict how well students are going to perform on a test by comparing their eye movements to the median eye movements.

Discussion

We found that eye movements during watching of instructional videos are similar between students, in particular if they are paying attention. The effect of attention is strong, allowing one to detect with a few minutes of gaze-position data whether the student is distracted. Consequently, and as predicted, we find that students performed well in subsequent quizzes if their eyes followed the material presented during the video in a stereotypical pattern. We replicated this finding in two subsequent laboratory experiments, where we confirmed that the effect persists when students do not expect to be quizzed and that the effect of attention does not depend on the specific type of video or the type of questions asked. The results also replicate in a classroom setting and in a large-scale online experiment with users at home using standard web cameras. By correlating with the median gaze positions one can avoid transmitting personal data over the internet. Thus, we conclude that one can detect

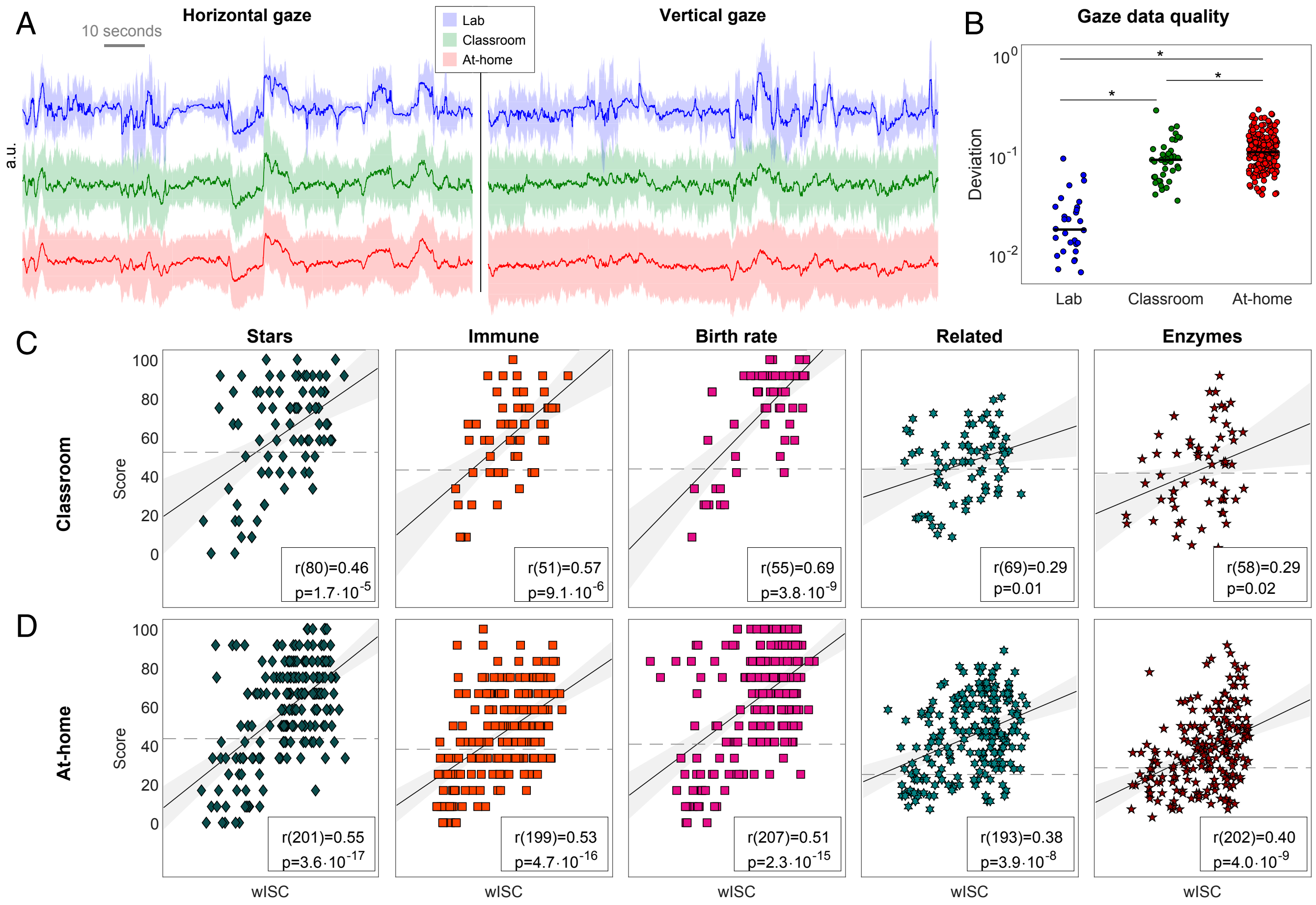

Fig. 3. Weighted intersubject correlation of eye movement measured using low-cost web camera predicts test scores. (*A*) Gaze position for "immune" video in laboratory, classroom, and at-home settings. Median and interquartile range are taken across subjects (solid line and shaded area, respectively). (*B*) Deviation of gaze position when subjects looked at four "validation" dots presented in sequence on the corners of the screen, collected in the laboratory, classroom, and at-home settings for the first video shown to subjects (*Materials and Methods*). *, a significant difference in means. (*C*) Eye-movement wISC is predictive of performance in the classroom. (*D*) Eye-movement wISC is predictive of performance in the at-home setting. See details in *SI Appendix*, Table S1.

students' attentional engagement during online education with readily available technology. In fact, we can predict how well students will perform on a test related to an instructional video, by looking at their eye movements while maintaining online privacy.

A traditional approach to assess overt visual attention is to simply determine whether subjects are looking at the stimulus, e.g., for video (20, 29). In contrast, ISC of eye movements measures whether subjects follow the visual content and not just whether they are looking at the video. Here we showed that ISC is a more faithful measure of active attentional engagement as it was a better predictor of test-taking performance compared to how long subjects looked at the video.

Note that the link between eye movements and test performance established here was purely correlational. It is possible that stronger students can both follow the video better and perform better in the test, without the need for a direct link between the two. It could also be that students with prior knowledge on the material were more interested and thus paid more attention. We built an analytic model assuming a common cause for intersubject correlation and test-taking performance. While we refer to this common cause as "attention," it really can refer to any internal state of a subject that may have a causal effect on test scores and eye movements such as alertness, interest, motivation, engagement, fatigue, etc. This causal model explained the data more accurately than a simple correlation. But ultimately, our study did not control attention prospectively and thus cannot conclusively answer the direction of this relationship.

We tested for recognition of factual information presented in the videos. Performance on these questions naturally depends on attention to the presentation of this factual information. For questions requiring comprehension, instead, it may be that students need time to think about material quietly without being absorbed by the video. Yet, we did not find a degradation in the ability to predict test scores from eye movement for the comprehension questions. However, a more nuanced analysis and larger sample size may be needed to establish a difference in our ability to predict comprehension vs. recognition performance.

ISC of eye movements varied significantly between subjects and videos. The variability between subjects is to a certain degree predictive of different test scores and thus we can ascribe it to genuine differences in attention. However, there is a significant variability in ISC across subjects even in the distracted condition, suggesting that baseline levels of ISC do vary between subjects, irrespective of attention. It is worth noting that we recruited a fairly diverse subject pool. In addition to the wide range in age and education levels reported here, the laboratory and classroom experiments likely also had a wide range of ethnicities, nationalities, and second language skills. The same is true for the online experiment, which recruited subjects through online platforms. This was a deliberate choice as we expected this to result in a wider range of attention levels.

ISC also differed significantly among videos. This again could be due to different levels of attention that the videos elicit, but it could also be due to differences in how the visual dynamic drives bottom–up attention; e.g., slower videos or less salient visuals may drive eye movements less vigorously (14, 20). In fact, we occasionally observed short segments of asynchronous eye movements, as gaze jumps back and forth between two prominent items in the video, but out of phase between subjects. Nevertheless, for the full videos ISC was always positive and correlated with learning gains. One caveat of using ISC is that its values may depend on spatial arrangements and the temporal dynamic. Therefore, ISC values should always be compared against a baseline that calibrates for differences between videos (30).

Eye tracking has been used extensively to study cognitive processing during reading (31) and visual search (32). In education research, specifically, eye tracking is often combined with a think-aloud protocol (33), whereby subjects verbalize their thinking while learning (34). This has been used to establish theories of learning (35). Eye-tracking data have also been used to classify different cognitive processes or to classify whether the viewer is an expert or naïve learner (36). But to our knowledge, eye movements have not been used to predict test-taking performance as we have done here.

Eye tracking has also been used in research specifically related to computer and online education. For instance, when learning from pictures and written text, fixation times and rereading predict learning performance (37). Showing the instructor's face while talking seems to help students attend to the material (38, 39), but there are mixed results on whether this is actually beneficial for learning (40, 41). These types of results required careful analysis of the exact content that is fixated upon. The method presented here assesses whether students are paying attention without the need for specific information about the contents of the video.

Given the link between the point of gaze and attention, attempts have been made to guide the attention of novice learners using cueing, e.g., pointing out where on a video or animation the student should pay attention. This can reduce cognitive load (42) and foster learning (43). There are also a few attempts of guiding attention of a novice by displaying the gaze position of an expert during problem solving (44) or during video presentation (45). However, this method has shown improved learning only in specific cases and requires careful manual annotations of the instructional videos (8).

Our analysis also included pupillary responses. That this should correlate between subjects is perhaps not surprising as it is strongly driven by luminance changes in the visual stimulus. The observation is that this correlation is modulated by attention. Interestingly, the ISC of pupil size remained predictive of test-taking performance even after regressing out luminance. Therefore, in the present study synchronization of pupil response is unlikely to result from luminance fluctuations and may be driven instead by other factors known to affect pupil size, such as arousal (39), cognitive effort (46), or attention (38). The present finding differs from the extensive literature on pupil size, which attempts to link pupillary response to specific events. For example, pupil size predicted reliably which stimuli were recalled, in particular for emotionally arousing stimuli (47). Pupil size has also been linked with cognitive effort, for instance, the effort associated with holding multiple items in working memory (46). In contrast to this traditional work on event-related pupil dilation we did not have to identify specific events in the stimulus. As with the eye movements, we can simply use other viewers as a reference to determine whether the pupil size is correlated and, if it is, anticipate high test scores. As online video communication becomes more prevalent, we expect future web cameras to have sufficient spatial resolution to capture pupil size. For the time being, we demonstrated current technology may suffice to predict learning performance based on eye movements alone.

Online education often struggles to persistently engage students' attention, which may be one of the causes for low retention (48). Students' online engagement is often measured in terms of the time spent watching videos (49), mouse clicks (50), or viewed content (51), and some important lessons have been learned from these outcome measures. For instance, videos should be short, be dynamic, and show the face of the instructor talking with enthusiasm (49). Our recent work has focused on measuring attentional engagement of the students by measuring their actual brain activity (52). Attempts to record brain signals in a classroom have been made (53–55), but typically require help from research personnel and may thus not be practical, particularly at home. The method we have presented here opens up the possibility to measure not just time spent with the material, but the actual engagement of the student's mind with the material, regardless of where the student is. With adequate data quality one may be able to even adapt the content in real time to the current attentional state of the student. In particular, for a synchronous online course, where students participate at the same time, real-time feedback to the teachers may allow them to adapt to students' level of attention in real time, much like real teachers in real classrooms. The internet has turned attention into a commodity. With video content increasing online, remote sensing of attention to video at scale may have applications beyond education, including entertainment, advertising, and politics. The applications are limitless.

Materials and Methods

Participants. A total of 1,182 subjects participated in one of five different experimental conditions. The first two experiments tested the learning scenario of online education, namely intentional learning (experiment 1, n = 27, 17 females, age 18 to 53 y, M = 26.74, SD = 8.98; 1 subject was removed due to bad data quality) and incidental learning (experiment 2, n = 31, 20 females, age range 18 to 50 y, M = 26.20, SD = 8.30 y; 3 subjects were removed due to bad signal quality). Experiment 3 was designed to investigate the effect of different video styles and types of questions (n = 30, 22 females, age 18 to 50 y, M = 25.73, SD = 8.85 y; 2 subjects were removed due to bad signal quality).

Participants for the laboratory experiments 1 to 3 were recruited from mailing lists of students at the City College of New York (CCNY) (n = 74) and the New York City section of craigslist.org (n = 14). n = 63 of 88 subjects self-reported to have some college education. CCNY has a diverse student body including age, ethnicity, second languages, skills, etc., and participants from the public in New York City only added to this diversity. Subjects were compensated for their time at a rate of $20 USD per hour, with the experiment lasting between 1 and 2 h.

Experiment 4 was designed to replicate the findings from the laboratory in a classroom setting. Participants were all second-year physics students enrolled in a common physics class at the City College of New York (n = 82, female = 21, age 18 to 40 y, M = 19.6, SD = 2.7 y). They were invited to participate at the beginning of one of their in-class lectures. For this experiment subjects were compensated with $3 USD per video and an additional $5 if they watched all five videos.

Experiment 5 replicated the finding from the laboratory in a home setting. Amazon Mechanical Turk and Prolific were used to recruit subjects (n = 1,012, 473 females, age range 18 to 64 y, M = 28.1, SD = 8.4 y). n = 532 of 1,012 subjects reported to have some college education and n = 353 had some high school education. Subjects of experiments 1 to 4 participated in only a single experiment; i.e., they were excluded from subsequent experiments. In experiment 5 subjects were allowed to participate in more than one assignment (video) so the total subject count is not unique subjects. For this experiment subjects were compensated with $3 USD per video.

The experimental protocol was approved by the Institutional Review Boards of the City University of New York. Documented informed consent was obtained from all subjects for laboratory experiments. Internet-based informed consent was given by subjects that were recruited for the online experiments.

Stimuli. The video stimuli are listed in *SI Appendix*, Table S2. Briefly, for experiments 1, 2, 4, and 5 we selected five videos from YouTube channels that post short informal instructional videos: "Kurzgesagt – In a Nutshell" and "Minute Physics." The videos cover topics relating to physics, biology, and computer science. Videos are short to match the limited attention span online (range, 2.4 to 6.5 min; average, 4.1 ± 2.0 min). Two of the videos ("Immune" and "Internet") used purely animations, where "Boys" used paper cutouts and handwriting. "Bulbs" and "Stars" showed a hand drawing illustrations aiding the narrative. For experiment 3, six video stimuli were selected using the following criteria: 1) The duration was limited to no more than 6 min (49) to ensure our subject would not lose interest (*SI Appendix*, Table S2; duration, 4.2 to 6 min; average, 5.15 ± 57 s). 2) The videos cover three different styles that are commonly found in large online educational channels on YouTube ("Khan Academy," "eHow," "Its ok to be smart," and "SciShow"). Here we use a nomenclature of video styles as found in the online instructional video community. "Mosquitoes" and "Related" were produced in the "Presenter and animation" style, which shows a presenter talking as pictures and animations are shown. "Planets" and "Enzymes" were produced in the "Presenter and glass board," which shows a presenter drawing illustrations and equations on a glass board facing the viewer. "Capacitors" and "Work energy" used the "Animation and writing hand" style, which shows a hand drawing animations. 3) Videos within each "style" cover two different topics each related to biology, astronomy, or physics. For experiments 4 and 5 a total of five videos were selected among the 11 videos used in experiments 1 to 3. Links to all videos are provided in *SI Appendix*, Table S2.

Procedure.

Laboratory experiments. In experiment 1 (intentional learning), subjects watched a video and answered afterward a short four-alternative forced-choice questionnaire. The five videos and question pairs were presented in random order. The subjects were aware that they would be tested on the material. The test covered factual information imparted during the video (11 to 12 recognition questions per video). Examples of questions and answer options can be found in *SI Appendix*, Table S2 and all can be found at https://osf.io/fjxaq/. In experiment 2 (incidental learning) subjects were not aware that they would be tested or asked questions regarding the material. They first watched all five videos and subsequently answered all of the questions (59 questions in total). In experiment 3, subjects were informed that questions regarding the material would be presented after each video and followed the procedure of experiment 1, using a different set of stimuli with six videos. The order of video presentation, questions, and answer options were randomized for all three experiments. Common for experiments 1 to 3, after subjects had watched all video stimuli and answered questions, they watched all of the videos again in a distracted condition using the same order as the attend condition. In this condition participants counted backward silently in the mind, from a randomly chosen prime number between 800 and 1,000, in decrements of 7. This task aimed to distract the subjects from the stimulus without requiring overt responses and is based on the serial subtraction task used to assess mental capacity and has previously been used to assess attention (7).

Online experiments. The web camera experiments (experiments 4 and 5) were carried out using Elicit, a framework developed for online experiments. In experiment 4 (classroom) students used the same computers they use for their class exercises. From the Elicit webpage subjects could select which video they wanted to watch from a list of five videos. Subjects were given a short verbal instruction besides the written instructions that were provided through the website. In experiment 5 (at-home) subjects could select human intelligence tasks (Amazon Mechanical Turk assignments) or assignments (Prolific) that contained a single video with questions and otherwise followed the same procedure as in experiment 4. For both experiments 4 and 5, subjects were informed that there would be questions regarding the material after the video. They first received instructions regarding the procedure, performed the webcam calibration to enable tracking of their eye movements, watched a single video, and answered a four-alternative choice questionnaire for that video. Subjects were allowed to perform more than one assignment, i.e., view more than one video and answer questions. In experiment 5 subjects were additionally shown a short instructional video on how to calibrate the webcam to track eye movements.

Online Eye Tracking Using Web Cameras. The webcam-based gaze position data were recorded using WebGazer (26). WebGazer runs locally on the subject's computer and uses the subject's webcam to compute the gaze position. The script fits a wireframe to the subject's face and captures images of the subject's eyes to compute where on the screen the subject is looking. Only the gaze position and the coordinates of the eye images used for the eye position computation were transmitted from the subject's computer to our web server. For the software to compute where on the screen the participant is looking, a standard nine-point calibration scheme was used. Subjects had to achieve a 70% accuracy to proceed in the experiment. Note that here we did transfer user data to the server for analysis. However, in a fully local implementation of the approach no user data would be transmitted. Instead, median eye positions of a previously recorded group would be transmitted to the remote location and median-to-subject correlation could be computed entirely locally.

Preprocessing of Webcam-Based Gaze Position Data. WebGazer estimates point of gaze on the screen as well as the position and size of the eyes on the webcam image. Eye position and size allowed us to estimate the movement of the subject in horizontal and vertical directions. The point of gaze and eye image position and size were up-sampled to a uniform 1,000 Hz, from the variable sampling rate of each remote webcam (typically in the range of 15 to 100 Hz). An inclusion criterion for the study was that the gaze position data should be sampled at least at an average of 15 Hz. Missing data were linearly interpolated and the gaze positions were denoised using a 300-ms-long median filter. Movements of the participant were linearly regressed out of the gaze position data using the estimated head position of the participant from the image patch coordinates. This was done since the estimated gaze position is sensitive to head movements (we found this regression increased the overall ISC). Subjects that had excessive movements were removed from the study (16 of 1,159 subjects; excessive movement is defined as 1,000 times the standard deviation of the recorded image patch coordinates in the horizontal, vertical, and depth directions). Blinks were detected as peaks in the vertical gaze position data after a 200-ms median filter. The onset and offset of each blink were identified as a minimum point in the first-order temporal derivative of the gaze position. Blinks were filled using linear interpolation in both the horizontal and vertical directions. Subjects that had more than 20% of data interpolated using this method were removed from the cohort (14 of 1,159 subjects). We could not compute the visual angle of gaze since no accurate estimate was available for the distance of the subject to the screen. Instead, gaze position is measured in units of pixels, i.e., where on the screen the subject is looking. Since the resolutions of computer screens vary across subjects, the recorded gaze position data in pixels were normalized to the width and height of the window the video was played in (between 0 and 1 indicating the edges of the video player). Events indicating end of the video stimuli ("stop event") were used to segment the gaze position data. The start time for each subject was estimated as the difference between the stop event and the actual duration of the video. This was done, since the time to load the YouTube player was variable across user platforms.

Estimating the Quality of Gaze Position. To compute the quality of the gaze position data, subjects were instructed to look at a sequence of four dots in each corner of the screen, embedded in the video stimuli before and after the video. The actual dot position on the individual screen was computed and compared to the captured eye gaze position of the WebGazer. The deviation was computed as the pooled deviation of the recorded gaze position from the position of the dot, while the subject looked at each dot. Poor data quality is indicated by higher deviation. Furthermore, subjects with low-quality calibration were identified by computing the spatial difference of recorded gaze position data of opposing dots in the horizontal and vertical directions when they were looking at the four dots. If the difference in recorded gaze position between dot pairs was in average negative, i.e., left/right reversed, the subject was excluded (135 of 1,159).

Preprocessing of Laboratory Gaze Position Data. In the laboratory (experiments 1 to 3) gaze position data were recorded using an Eyelink 1000 eye tracker (SR Research Ltd.) at a sampling frequency of 500 Hz using a 35-mm lens. The subjects were free to move their heads, to ensure comfort (no chin rest). A standard nine-point calibration scheme was used, using manual verification. To ensure stable pupil size recordings, the background color of the calibration screen and all instructions presented to the subjects were set to be the average luminance of all of the videos presented during the experiment. In between each stimulus presentation a drift check was performed and tracking was recalibrated if the visual angular error was greater than 2°. Blinks were detected using the SR research blink detection algorithm and remaining peaks were found using a peak picking algorithm. The blink and 100 ms before and after were filled with linearly interpolated values.

Intersubject Correlation and Attention Analysis of Gaze Position Data. Intersubject correlation of eye movements was calculated by 1) computing the Pearson's correlation coefficient between a single subject's gaze position in the vertical direction with that of all other subjects while they watched a video; 2) obtaining a single ISC value for a subject by averaging the correlation values between that subject and all other subjects (ISC); and 3) then repeating steps 1 and 2 for all subjects, resulting in a single ISC value for each subject. We repeat these three steps for the horizontal eye movements $\mathrm{ISC}_{horizontal}$ and the pupil size ISC_{pupil}. To obtain the measure used for the laboratory experiments we averaged the three ISC values which we call $\mathrm{ISC} = (\mathrm{ISC}_{vertical} + \mathrm{ISC}_{horizontal} + \mathrm{ISC}_{pupil})/3$. The ISC values for the attend and distract conditions were computed on the data for the two conditions separately. To test whether ISC varies between the attend and distract conditions, a three-way repeated-measures ANOVA was used with fixed effect of video and attentional state (attend vs. distract) and random effect of subject. As an additional measure the receiver operating characteristic curve (ROC) was used. Each point on the curve is a single subject. To quantify the overall ability of ISC to discriminate between attend and distract conditions the area under the ROC curve (AUC) is used. To test for the effect of motivation, ISC was computed for each video in the attend condition and averaged across all videos. Since the distribution was not Gaussian, we tested for a difference in median ISC values with a Wilcoxon rank sum test. To test for the effect of video style on the attentional modulation of ISC we performed a three-way repeated-measures ANOVA. The random effect was subject and fixed effects were stimuli, attentional condition, and video style.

Weighted Intersubject Correlation of Eye Movements. For the experiments with the web camera in the classroom and at home we compute for each time point in the video the median gaze position across all subjects (Fig. 3*A*). We then compute the Pearson's correlation coefficient of that median time course with the time course of gaze position of each subject. We refer to this as median-to-subject correlation, $\mathrm{MSC}_{vertical}$ and $\mathrm{MSC}_{horizontal}$. Note that in principle this can be computed with the median gaze positions previously collected on a sample group for each video. To compute this remotely without transmitting the gaze data of individual users, one would transmit the median gaze positions to the remote user of the online platform (two values for each time point in the video). MSC can then be computed locally by the remote user. Eye velocity has been demonstrated to be a useful measure of synchronous eye movements (16). We therefore compute in addition MSC for the velocity of eye movements as follows. First, we compute movement velocity by taking the temporal derivative of horizontal and vertical gaze positions using the Hilbert transform. We form two-dimensional spatial vectors of these velocity estimates (combining Hilbert transforms of horizontal and vertical directions). These vectors are normalized to unit length. The median gaze velocity vector is obtained as the median of the two coordinates across all subjects. The median-to-subject correlation of velocity, $\mathrm{MSC}_{velocity}$, is then computed as the cosine distance between the velocity vectors of each subject and the median velocity vector, averaged over time. Finally, we combine the three MSC measures to obtain a single weighted intersubject correlation value for each subject, $\mathrm{wISC} = w_1\mathrm{MSC}_{vertical} + w_2\mathrm{MSC}_{horizontal} + w_3\mathrm{MSC}_{velocity}$, following our previous work on ISC of neural signals (27, 28). The weights w_i are chosen to best predict test scores with the constraint that they must sum up to 1 and that they are all positive. This is done with conventional constrained optimization. The constraints ensure that the wISC values are bounded between -1 and 1. To avoid a biased estimate of predictability we optimize these weights for each subject on the gaze/score data leaving out that subject from the optimization; i.e., we use leave-one-out cross-validation.

Frequency-Resolved Analysis of ISC. We performed a frequency analysis to investigate at which time scale eye movements and pupil size synchronize between subjects. The vertical and horizontal gaze position signal was bandpass filtered using fifth-order Butterworth filters with logarithmic spaced center frequencies with a bandwidth of 0.2 of the center frequency. The ISC was computed for each subject in each frequency band (experiment 2 on all five videos). To obtain a single ISC value per frequency band we average ISC values for all videos, for all subjects, and across the two directions (horizontal and vertical). The gray-shaded intervals around the mean values (Fig. 1*D*) are the standard error across subjects. The same analysis was done on the pupil size.

Student Learning Assessment. Four-alternative, forced-choice questions were used to assess the performance of students (score). Test performance was calculated as the percentage of correct responses each student gave for each video. For questions that had multiple correct options, points were given per correct selected option and subtracted per incorrect selected option. The questionnaires were designed in pilot experiments to yield an even distribution of answer options from subjects that had not seen the videos. All questions and answer options can be found at https://osf.io/fjxaq/.

To estimate the baseline difficulty of the questions, separate naïve cohorts of subjects were given the same questions without seeing the videos. Two different cohorts were recruited from the City College of New York to compare against the cohorts recruited for experiments 1 to 4 (experiments 1, 2, and 4, n = 26; experiment 3, n = 15) and a third from Prolific to compare against the at-home experiment cohort (experiment 5, n = 25).

All questions were categorized as either recognition or comprehension questions following Bloom's taxonomy (56) and defined here specifically as follows: Recognition–Question can be answered by remembering a word, phrase, or number which was specifically stated in the video and does not require understanding of scientific concepts to answer correctly; Comprehension–Question involves an application or interpretation of ideas presented in the video or identification of concepts developed in the video that likely requires understanding to be able to answer correctly. To decide on the category for each question, we independently rated each question and the majority rating was selected as the final categorization (see ratings at https://osf.io/fjxaq/).

Relating Student Test Performance and ISC. When evaluating the different learning scenarios (incidental and intentional learning) in experiments 1 and 2, students' scores and ISC values were averaged across all videos. ISC was compared to student test performance by computing the Pearson's correlation coefficient between ISC and test performance. Similarly, to test the effect of video style, the ISC and scores for each subject were averages for the videos produced in different styles and correlated using Pearson's correlation. Testing the connection between ISC and test scores on each individual video, subjects' scores were compared with the ISC using Pearson's correlation. To test whether there is a significant difference in correlation between comprehension vs. recognition questions and ISC we used the same ISC values and performed a test between correlation values with a shared dependent variable (57). Testing how well eye-movement ISC can predict the performance of students on tests regarding the material in the online setting, we use leave-one-out cross-validation. We estimate the attention model (see *SI Appendix*, section S4 for description) on all subjects leaving out one subject's ISC values and their corresponding test scores. We then estimate how well ISC predicts the test score on the left-out subject. We do this for all subjects and compute the median absolute deviation between the prediction and the actual score. To test whether our eye-movement ISC model is statistically better than a naïve model (only predicting the average score), we subtract the prediction errors of the two models and perform a two-sided sign test.

Data Availability. Anonymized data to produce each figure in matlab format is available in the Open Science Framework (https://osf.io/m7gj4/). A full list of questions and answer options can be found in Open Science Framework (https://osf.io/fjxaq/). The code used to carry out the online experiment is available in Github (https://github.com/elicit-experiment).

ACKNOWLEDGMENTS. We thank Iain Bryson from Freeroaming Solutions for his invaluable help in developing Elicit, which allowed us to carry out the online experiments. Without him this would not have been possible. We acknowledge the help of James Hedberg for supporting collection of data from the physics classroom cohort. We thank Uri Hasson, Ofer Tchernichowski, and Hernan Makse for helpful comments on an early version of this paper. We acknowledge NSF Grant DRL-1660548 for supporting this project.

1. R. G. Packard, The control of "classroom attention": A group contingency for complex behavior 1. *J. Appl. Behav. Anal.* **3**, 13–28 (1970).
2. D. M. Bunce, E. A. Flens, K. Y. Neiles, How long can students pay attention in class? A study of student attention decline using clickers. *J. Chem. Educ.* **87**, 1438–1443 (2010).
3. H. Deubel, W. X. Schneider, Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vis. Res.* **36**, 1827–1837 (1996).
4. J. E. Hoffman, B. Subramaniam, The role of visual attention in saccadic eye movements. *Percept. Psychophys.* **57**, 787–795 (1995).

5. C. E. Wolff, N. van den Bogert, H. Jarodzka, H. P. A. Boshuizen, Keeping an eye on learning: Differences between expert and novice teachers' representations of classroom management events. *J. Teach. Educ.* **66**, 68–85 (2015).
6. H. J. Bucher, P. Schumacher, The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. *Communications* **31**, 347–368 (2006).
7. L. Lorigo *et al.*, Eye tracking and online search: Lessons learned and challenges ahead. *J. Am. Soc. Inf. Sci. Technol.* **59**, 1041–1052 (2008).
8. H. Jarodzka, K. Holmqvist, H. Gruber, Eye tracking in educational science: Theoretical frameworks and research agendas. *J. Eye Mov. Res.*, 10.16910/jemr.10.1.3 (2017).
9. D. A. Slykhuis, E. N. Wiebe, L. A. Annetta, Eye-tracking students' attention to PowerPoint photographs in a science education setting. *J. Sci. Educ. Technol.* **14**, 509–520 (2005).
10. F. Y. Yang, C. Y. Chang, W. R. Chien, Y. T. Chien, Y. H. Tseng, Tracking learners' visual attention during a multimedia presentation in a real classroom. *Comput. Educ.* **62**, 208–220 (2013).
11. U. Hasson, E. Yang, I. Vallines, D. J. Heeger, N. Rubin, A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* **28**, 2539–2550 (2008).
12. J. M. Franchak, D. J. Heeger, U. Hasson, K. E. Adolph, Free viewing gaze behavior in infants and adults. *Infancy* **21**, 262–287 (2016).
13. U. Hasson *et al.*, Neurocinematics: The neuroscience of film. *Projections* **2**, 1–26 (2008).
14. M. Dorr, T. Martinetz, K. R. Gegenfurtner, E. Barth, Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* **10**, 28 (2010).
15. C. Christoforou, S. Christou-Champi, F. Constantinidou, M. Theodorou, From the eyes and the heart: A novel eye-gaze metric that predicts video preferences of a large audience. *Front. Psychol.* **6**, 579 (2015).
16. K. Burleson-Lesser, F. Morone, P. DeGuzman, L. C. Parra, H. A. Makse, Collective behavior in video viewing: A thermodynamic analysis of gaze position. *PloS One* **12**, e0168995 (2017).
17. T. J. Smith, P. K. Mital, Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *J. Vis.* **13**, 16 (2013).
18. M. L. Lai *et al.*, A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educ. Res. Rev.* **10**, 90–115 (2013).
19. H. X. Wang, J. Freeman, E. P. Merriam, U. Hasson, D. J. Heeger, Temporal eye movement strategies during naturalistic viewing. *J. Vis.* **12**, 16 (2012).
20. L. Itti, P. Baldi, Bayesian surprise attracts human attention. *Vis. Res.* **49**, 1295–1306 (2009).
21. M. Hegarty, S. Kriz, C. Cate, The roles of mental animations and external animations in understanding mechanical systems. *Cogn. InStruct.* **21**, 209–249 (2003).
22. R. E. Mayer, M. Hegarty, S. Mayer, J. Campbell, When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *J. Exp. Psychol. Appl.* **11**, 256–265 (2005).
23. G. R. Loftus, Eye fixations and recognition memory for pictures. *Cogn. Psychol.* **3**, 525–551 (1972).
24. M. C. Potter, E. I. Levy, Recognition memory for a rapid sequence of pictures. *J. Exp. Psychol.* **81**, 10–15 (1969).
25. F. W. Schneider, B. L. Kintz, An analysis of the incidental-intentional learning dichotomy. *J. Exp. Psychol.* **73**, 85–90 (1967).
26. A. Papoutsaki *et al.*, "Webgazer: Scalable webcam eye tracking using user interactions" in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, S. Kambhampati, Ed. (AAAI Press, 2016), pp. 3839–3845.
27. J. P. Dmochowski *et al.*, Audience preferences are predicted by temporal reliability of neural processing. *Nat. Commun.* **5**, 4567 (2014).
28. S. S. Cohen, S. Henin, L. C. Parra, Engaging narratives evoke similar neural activity and lead to similar time perception. *Sci. Rep.* **7**, 4578 (2017).
29. O. Le Meur, A. Ninassi, P. Le Callet, D. Barba, Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric. *Signal Process. Image Commun.* **25**, 547–558 (2010).
30. J. J. Ki, S. P. Kelly, L. C. Parra, Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *J. Neurosci.* **36**, 3092–3101 (2016).
31. M. A. Just, P. A. Carpenter, A theory of reading: From eye fixations to comprehension. *Psychol. Rev.* **87**, 329–354 (1980).
32. K. Rayner, The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* **62**, 1457–1506 (2009).
33. J. Preissle, M. D. Le Compte, *Ethnography and Qualitative Design in Educational Research* (Academic Press, 1984).
34. T. Van Gog, F. Paas, J. J. Van Merriënboer, Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Appl. Cogn. Psychol.* **19**, 205–221 (2005).
35. M. van Someren, Y. Barnard, J. Sandberg, *The Think Aloud Method: A Practical Approach to Modeling Cognitive Processes* (Academic Press, 1994).
36. S. Eivazi, R. Bednarik, "Predicting problem-solving behavior and performance levels from visual attention data" in *Proceedings of the Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI* (Association for Computing Machinery, New York, NY, 2011), pp. 9–16.
37. S. C. Chen *et al.*, Eye movements predict students' computer-based assessment performance of physics concepts in different presentation modalities. *Comput. Educ.* **74**, 61–72 (2014).
38. S. Mathôt, S. Van der Stigchel, New light on the mind's eye: The pupillary light response as active vision. *Curr. Dir. Psychol. Sci.* **24**, 374–378 (2015).
39. M. M. Bradley, L. Miccoli, M. A. Escrig, P. J. Lang, The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**, 602–607 (2008).
40. L. Fiorella, A. T. Stull, S. Kuhlmann, R. E. Mayer, Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *J. Educ. Psychol.* **111**, 1162 (2019).
41. M. van Wermeskerken, S. Ravensbergen, T. van Gog, Effects of instructor presence in video modeling examples on attention and learning. *Comput. Hum. Behav.* **89**, 430–438 (2018).
42. B. B. De Koning, H. K. Tabbers, R. M. Rikers, F. Paas, Attention guidance in learning from a complex animation: Seeing is understanding?. *Learn. InStruct.* **20**, 111–122 (2010).
43. B. B. De Koning, H. K. Tabbers, R. M. Rikers, F. Paas, Attention cueing as a means to enhance learning from an animation. *Appl. Cognit. Psychol.* **21**, 731–746 (2007).
44. T. Van Gog, H. Jarodzka, K. Scheiter, P. Gerjets, F. Paas, Attention guidance during example study via the model's eye movements. *Comput. Hum. Behav.* **25**, 785–791 (2009).
45. H. Jarodzka, K. Scheiter, P. Gerjets, T. Van Gog, In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learn. InStruct.* **20**, 146–154 (2010).
46. T. Piquado, D. Isaacowitz, A. Wingfield, Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* **47**, 560–569 (2010).
47. A. Bergt, A. E. Urai, T. H. Donner, L. Schwabe, Reading memory formation from the eyes. *Eur. J. Neurosci.* **47**, 1525–1533 (2018).
48. S. I. de Freitas, J. Morgan, D. Gibson, Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. *Br. J. Educ. Technol.* **46**, 455–471 (2015).
49. P. J. Guo, J. Kim, R. Rubin, "How video production affects student engagement: An empirical study of MOOC videos" in *Proceedings of the First ACM Conference on Learning @ Scale Conference, L@S '14* (Association for Computing Machinery, New York, NY, 2014), pp. 41–50.
50. M. Ginda, M. C. Richey, M. Cousino, K. Börner, Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design. *PLoS One* **14**, e0215964 (2019).
51. D. Lagun, M. Lalmas, "Understanding user attention and engagement in online news reading" in *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* (Association for Computing Machinery, New York, NY, 2016), pp. 113–122.
52. S. S. Cohen, J. Madsen *et al.*, Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiol. Learn. Mem.* **155**, 60–64 (2018).
53. S. Dikker *et al.*, Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr. Biol.* **27**, 1375–1380 (2017).
54. A. T. Poulsen, S. Kamronn, J. Dmochowski, L. C. Parra, L. K. Hansen, EEG in the classroom: Synchronised neural recordings during video presentation. *Sci. Rep.* **7**, 43916 (2017).
55. D. Bevilacqua *et al.*, Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: Evidence from a real-world classroom electroencephalography study. *J. Cogn. Neurosci.* **31**, 401–411 (2019).
56. B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, D. R. Krathwohl, *Taxonomy of Educational Objectives: Cognitive Domain* (Longman Group, 1956), vol. 1.
57. J. H. Steiger, Tests for comparing elements of a correlation matrix. *Psychol. Bull.* **87**, 245–251 (1980).