Sleepers' Lag: Study on Motion and Attention

Mirko Raca

CHILI Laboratory

École polytechnique fédérale de Lausanne, Switzerland

mirko.raca@epfl.ch

Roland Tormey

CAPE

École polytechnique fédérale de Lausanne, Switzerland

Pierre Dillenbourg

CHILI Laboratory

École polytechnique fédérale de Lausanne, Switzerland

ABSTRACT: Body language is an essential source of information in everyday communication. Low signal-to-noise ratio prevents us from using it in the automatic processing of student behaviour, an obstacle that we are slowly overcoming with advanced statistical methods. Instead of profiling individual behaviour of students in the classroom, the idea is to compare students and connect the observed traits to different levels of attention. With the usage of novel techniques from the field of computer vision, we focus on features that can be automatically extracted with a system of cameras, by means of passive observation of the classroom population. We show parallels between our work and previous theories and formulate a new concept for measuring the level of attention based on synchronization of student body movement. We observed that students with lower levels of attention are slower to react than focused students, a phenomenon we named "sleepers' lag." This realization may give rise to novel measurements that can act as a technological support for teacher metacognition. The goal is to improve the teacher–student conversation and to propose techniques that can enable a shorter feedback loop of the teacher's performance compared to the current-day methods.

Keywords: Video analysis, computer vision, tracking, body motion, classroom, interpersonal synchronization, orchestration

1   INTRODUCTION

Attention is the "gateway" through which students learn (Shell et al., 2010), but this essential trait is easy to lose and hard to assess. So, how can the lecturer "measure" the attention of students during the class? Typically, classroom interactions (Q-A, interactions, demonstrations) are used as proxies, but in standard lecture settings, student participation is very low. Teacher observations tend to be based on a small sample of high-interaction individuals, while fewer than 40% of students actively engage in the conversation (Howard & Henney, 1998).

Our approach is based on an attempt to formalize observation of teacher effect on students. Coe, Aloisi, Higgins, and Major (2014) confirmed the validity of classroom observation as a method of teacher assessment. But when the operation is carried out by individuals assessing the teacher, Bernstein (2008) noted that the quality of the process is largely dependent on the training of the observers. Another constraint that Range, Duncan, and Hvidston (2013) noted was the time limit for the post-observational conference, which should be within five days of the intervention.

The direct problem — evaluating the attention that the teacher's action will attract, is highly subjective and impossible to automatize. Modern approaches attempt to classify appealing body language and presentation styles, but in order to assess the effect of the approach, we need to turn to the audience.

We consider teacher performance and student understanding as two sides of the same coin. As Hattie's (2013) meta-analysis noted, presenting the effect of their intervention back to the teachers is one of the strongest effects among educational interventions. In order to re-connect the two sides of the classroom into a mutually beneficial conversation, we aim to present a technology that can provide teachers with seamless feedback. Timperley, Wilson, Barrar, and Fung (2008) describe a broader set of principles as a "knowledge-building cycle" — a set of efforts needed to continue teachers' professional improvement. The methods in our approach are already well established in the human conversation as the grounding principle (Clark & Brennan, 1991) and the back channel (Vinciarelli et al., 2012).

Our technological intervention is aimed towards amplifying the back channel, and focusing the teacher towards classroom interactions. It is important to note our intention to augment (expand) the feedback loop, and not to replace the power of the teacher's own observation. Even though current challenges in the technological domain lie in achieving human performance, excluding the teacher from the learning loop would be a mistake; as the orchestrator of the learning process (Dillenbourg & Jermann, 2010), the teacher is responsible for integrating the information into the overall learning experience.

Without overloading the students with gadgets and formally structured procedures that dictate the format of the learning experience, we aim to implement our system with a set of cameras. The base for our observations is human activity in its most basic form — movement.

In this paper, we present the method for measuring movement in a classroom and the procedure used to relate the gathered information to students' subjective perceptions of their own attention. The main contribution is the concept of measuring the speed of student reactions in class to detect students with lower attention. The concept is based on the idea that students focused on the lecture would react in the moment to the important information being presented, while distracted students would be slower to note it. This is the concept we call "sleepers' lag." The higher the variance in reaction time to the common stimuli (in our case to the teacher's presentation) — the lower the attention of the classroom audience.

Our other conclusions go further into exploring how the geometry of the classroom and immediate surroundings affect the individual student. This sets the ground for "student-centred" observation of the classroom, as opposed to the dominant trend of exploration that considers the teacher as the only stimulus present.

2   RELATED WORK

Traditional classrooms (in both talk format and seating configuration) remain the dominant format of lecturing on all levels of formal education today (Moore, 1989). There have been many critiques of the format, noting that the classroom's geographical configuration makes it difficult to develop the teacher–student relationship and understanding beyond stereotypes (Hargreaves, 2000). And while some claim that the current organizational setup evolved for practical reasons (Koneya, 1976) we cannot ignore the difficulties that teacher face in keeping student attention over time (Middendorf & Kalish, 1996; Wilson & Korn, 2007) and on-task (Rosengrant et al., 2012).

The set of theories we group under the name "teacher-centric" focus on the teacher and the teacher's impact on the classroom. As the primary orchestrator of the learning process (Dillenbourg et al., 2011; Dillenbourg & Jermann, 2010), teachers take on the responsibility that begins with educational presentation, follows through pedagogical guidance (Corcoran & Tormey, 2012), and hopes to achieve students' personal transformation (Whitcomb, Borko, & Liston, 2008). The teacher's role in the classroom has been characterized as emotional labour (Hargreaves, 2000) and cognitively demanding (Emmer & Stough, 2001). In many instances, a good teacher is characterized by the ability to present the teaching material in a way that engages students, this being the major difference between a novice and an experienced teacher (Borko & Livingston, 1989), confirming the need for the teacher to be a reflective practitioner (Schön, 1983).

The geometry of the classroom can also be an emotional barrier for more natural interaction (Hargreaves, 2000). Students in the front rows are perceived as being "more interested" (Daly & Suite, 1981). The bulk of communication is oriented in a T-shaped region with the highest concentration of interaction focused on the front and centre of the classroom (Adams, 1969). This not only affects the teacher's perception, but students also adjust to the geometry of the classroom, with those seeking interaction tending to sit in the high-interaction zone (Altman & Lett, 1970). The seating arrangement also amplifies student interactions — making high-verbalizers more active in the high-interaction zone, and low-verbalizers even less active in the low-interaction zone (the edges of the classroom) (Koneya, 1976). The classroom environment greatly affects the perceptions of teacher and students, but this does not always work in favour of the learning process.

Being far away from the teacher goes beyond just teacher perception. On the "student-centric" side of research, Daum (1972) found that distance from the teacher also has a significant effect on the success of students. Finn, Pannozzo, and Achilles (2003) found that smaller class sizes (fewer than 15 students) affect the quality of the lecture in two ways — the teachers take less time to manage the learning process, but more importantly student-to-student interaction also improved. As students grow up in the school system, the relationship between teacher and student becomes less emotionally involved (Hargreaves, 2000) and their participation in classroom activity decreases (Marks, 2000). This seems closely related to students becoming more accustomed to studying in a large groups, where individual visibility is uncertain (Finn et al., 2003) and circumstances allow for easy diffusion of responsibility and social loafing (Forsyth, 2009). It

is common for students to have more practical goals (i.e., good grades) than purely academic growth (Allen, 1986).

Irrespective of position or grades, students have difficulty maintaining their attention throughout the duration of a lecture (Rosengrant et al., 2012). Attention "can be partially defined as the selection, activation, and maintenance of mental focus on some stimuli (external or internal) accompanied by the blocking of other stimuli" (Rapp, 2006). Roda and Thomas (2006) noted it as our biological defense against informational overload coming primarily from the external environment. Even if it is not clearly quantifiable how long it takes students to "zone out" during a lecture, proposed measurements of between 10 minutes (Wilson & Korn, 2007) and 20 minutes (Middendorf & Kalish, 1996) are far less than the average duration of a lecture. Moore (1989) recognized that student attention is divided between three types of interactions: i) learner–content, ii) learner–instructor, and iii) learner–learner, in which the second type has priority over the other two in class, due to its limited availability. Roda and Thomas (2006) produced a detailed specification of how attention should be handled in the domain of human–computer interaction, but outside of this strictly technical domain, the rules become less defined. Various approaches to determine user attention were formulated with eye-tracking research being the prevalent method for its measurability (Nüssli, 2011). Head-pose was also found to be a good indicator of visual attention with 88% accuracy (Stiefelhagen & Zhu, 2002). With the goal of raising the accuracy of prediction, other methods introduced various complementary measurements, such as EEG devices and heart-rate monitors (Chen & Vertegaal, 2004) and other contextual information (Arroyo et al., 2009; El Kaliouby & Robinson, 2004; Horvitz, Kadie, Paek, & Hovel, 2003) with the constant trade-off between the complexity of the measuring apparatus and the confidence of the prediction. In the area of measuring "expertise," focusing solely on the activity as the cue, successful attempts at observing different behavioural patterns have been observed in both expert and novice categories (Worsley & Blikstein, 2013).

In a broader scope, Social Signal Processing (SSP) research field (Vinciarelli, Pantic, & Bourlard, 2009) poses the idea that machine interpretation of simple human actions has reached its limit, and in order to improve automatic analysis, we need to encode social context (Vinciarelli et al., 2012). With scope well beyond the classroom, attempts have already been made in interpreting the behaviours of large groups at sporting events (Conigliaro, Setti, Bassetti, Ferrario, & Cristani, 2013) and in public spaces in general (Bazzani et al., 2013). The first results showed promise, but with the crudeness of the initial findings, we are again reminded of the complexity of human interaction. Gatica-Perez (2009) showed the need for identifying this new branch of research, as papers on the topic are currently distributed over several scientific domains based on the methods, applications, etc.

Our research aims to scaffold teacher's perception of the students and raise awareness about student reception of the lecture. Some of the current methods of doing so are focused on the web-domain interactions (Dyckhoff, Lukarov, Muslim, Chatti, & Schroeder, 2013), feedback devices such as clickers (Caldwell, 2007), or mobile phone apps (Rivera-Pelayo, Munk, Zacharias, & Braun, 2013). We looked to the research on unobtrusive measurements (Webb, Campbell, Schwartz, & Sechrest, 1999) in order not to disturb the classroom ecosystem. The topic is sensitive because, as Heylighen (2002) has already noted, information overload leads to such dangerous pitfalls as anxiety, stress, and alienation. In the midst of

such a mentally demanding task as teaching, we must be careful when introducing new elements since the main bottleneck may still remain in the teacher's head. We took important cues from ubiquitous computing principles (Weiser, 1991), and interventions in which the information was available when needed, but was not the focus of the activity (Bachour, 2010; Alavi, 2012).

3 THEORETICAL BACKGROUND

In everyday communication, grounding occurs seamlessly throughout the conversation. The work of Clark and Brennan (1991) defines grounding as the collective process by which participants try to establish the mutual belief that all sides understand each other in order to continue the conversation successfully. In one-on-one communication, grounding is essential and completely interwoven with other activities; in the classroom, however, the feedback component is much weaker. Lecturing is inherently imbalanced between the two grounding phases — i) presentation, and ii) acceptance of information — largely in favour of the teacher.

The "acceptance phase" is well developed in the educational domain, out of the need to formalize the process. The evaluation of student knowledge takes many forms, and in order to show how teachers use different types of evaluation, we emphasize the following properties:

- **Social scope**: Differentiating between the evaluation of a single person, work-group, class, generation, etc.

- **Delay**: Time between the presented information and proof of its assimilation. While conversational grounding happens instantaneously, more formal techniques have longer delays, either within one work-unit (question and answer pair during class), or several days (quiz results, final exam, etc.).

- **Confidence**: We can never be certain if the presented knowledge displays actual comprehension, but different methods offer results of higher or lower reliability. While students nodding can be no more than a minimal-effort conversational continuer, a fully answered open-ended question in the final quiz will be a more reliable indicator of actual understanding.

- **Material scope**: Depending on the formulation of the question, the answer might require the student to demonstrate knowledge of a single definition, explain material presented within the lesson (topic), or connect several scientific areas.

While the formal education process requires wide material and social scope, there is little space for intervention and correction of student knowledge. In order to do preventive evaluation, teachers often use a smaller material scope and shorter delay (e.g., continuous testing). In doing so, low performance can be explained by "did not study hard enough" instead of "material was not appropriately presented." Due to the large number of factors influencing learning, the longer the delay — the higher the distribution of responsibility.

The value of feedback to teachers has been proven highly effective. In an exhaustive meta-study on the effects of different factors on learning, Hattie (2013) placed feedback to teachers as the tenth most influential factor analyzed in terms of student success. But current systems for teaching evaluation are typically carried out at the end of term, which effectively dissociates the student grade from any single action on the part of the teacher. In terms of deliberate practice, Ericsson (2008) suggests that the "best training situations focus on activities of short duration with opportunities for immediate feedback, reflection, and correction." But what does this mean for the feedback loop to the teacher as the performer of teaching activities?

To perform spontaneous self-evaluation, teachers are reduced to the conversational check-in with the class, which offers short delay and low material scope, but also low social scope and confidence. We address each point separately.

Low **material scope** means frequent requests for feedback from students, which can be automatically carried out by maintaining eye contact. This "focused attention" on the individual student is used both as a feedback device and as a method of reconnecting the absent-minded student to the classroom material.

Low **social scope** comes purely from our mental constraints. Confronted with a group of people, a human observer is sequentially analyzing each individual. Again, to widen the scope of the analysis, the teacher would need to spend more time evaluating. A potential way around this bottleneck is to generalize or extrapolate information about the student state, which we will address shortly.

We can assume that **low confidence** is caused in part by conversational conformity and peer pressure. In the brief interaction with the teacher, a student engaged directly in the conversation is often tricked into simulating positive grounding evidence by providing a minimal-effort "continuer" — such as a head nod (Clark & Brennan, 1991) — motivated primarily by the need to continue the lecture (effectively a "conversation" between teacher and student). A secondary obstacle for reporting actual understanding of the lesson is peer-pressure and conformity, which implicate that the student needs to step away from the anonymity of the classroom (Forsyth, 2009) and admit a lack of understanding publicly. The source of both problems is that the feedback requires direct and intentional interaction with the teacher. The "intentionality" of feedback is common in most other approaches, and the main issue we overcome with the observational approach.

In order to keep up with the teaching schedule, teachers have several generalization tools that they can use to infer attention and comprehension:

- **Teacher experience:** Developing intuition about student reactions is the slowest method to train. This automation of thinking and mental shortcuts (Kahneman, 2011) is usually found in more experienced teachers. Unfortunately, due to the slow feedback loop, this can be also the most erroneous method (Ericsson, 2008).

- **Familiarity with the student in question:** Built through the lens of experience, but shorter in the

time-scope, familiarity with individual students can provide useful feedback. The main difference lies in the locus of information — while the experience is primarily associated with the actions of the teacher, main source of information remains the individual student.

- **How the lecture is going so far:** This is a short-term temporal method, in which no individual is being considered, but rather the overall reception of the lesson by the class.

- **Attitude in the classroom: The social dimension.** Even if some students are not visible, the teacher can infer "general acceptance" of the material by "reading the audience" as actors do on stage.

Dominant dimensions that overlap in the noted methods include i) experience, ii) time and iii) the social dimension. Given that each method interpolates these three components to different degrees, we base our approach primarily on the socio-temporal dimensions, in service of scaffolding the third component, which remains connected to the teachers themselves. This naturally assigns the approach with attributes such as wide social scope and independence of the material scope — given that the automated measurements can be applied at any time. The approach attempts to access the socially visible information into which we have limited access due to our biological limitations, amplifying the back-channel communication. Previous work stated that body language, while rich in semantics, is low on syntax (Vinciarelli et al., 2012) — which makes it implicitly unreliable. But the availability of data emitted from the students as informative (carries meaning) if not communicative (not purposefully used for communication) signals provides fertile ground for analysis.

3.1   Theoretical Assumptions

Our initial hypothesis for the experiment was that we could detect consistent groups of students by common behaviour patterns. An example of consistency would be a group of students listening to the lecture versus students looking out the window. Second hypothesis was that people in the visible surroundings of an individual affect that person (student) by their non-verbal cues. We considered body language in its most basic form and compared the co-occurrences of motion (*co-movement*) between pairs of students. We also related our observations to students' levels of attention.

From the dual eye-tracking theory, we know that the quality of collaboration (Richardson, Dale, & Kirkham, 2007) and understanding (Jermann & Nüssli, 2012) between two persons can be assessed by analyzing the consistency of their gaze patterns. We draw an analogy with these conclusions in the domain of motion in the classroom, with the hypothesis that students who listen to the teacher will be more likely to move in a synchronized manner, while an absent-minded student will act on his/her own internal rhythm. Synchronized motion is not limited to any specific action, but can be explained using the example of taking notes — attentive students would turn the pages on the handouts and note important facts as they are presented in class. More than a reaction to the lecture's audio/visual stimulus, motion can be seen as an agreement of the audience. If the students agree that an event outside the classroom (e.g., loud noise, truck) is more important than the lecture, they would still have a synchronized motion (everybody looking out the window) but caused by a different stimulus than the teacher.

Synchronization in the class was studied in a dyadic fashion, by comparing each pair of students. Depending on the relative location between the two students considered in the pair, we divided the dyads into three conditions based on their mutual visibility (as described in Section 247.2).

Given that learning is not a strictly formalized activity, reactions of students can vary or be completely blank. In dual eye tracking, a delay of 2 seconds between the speaker's and the listener's gaze during the moments of referencing has been identified (Richardson et al., 2007), with the conclusion that the comprehension between participants is inversely proportional to the time lag. Based on this, we define two movements as co-movement if it happens within ±4 seconds from each other (depicted in Figure 1a). We differentiate between i) **perfect synchronization** (<2sec apart), ii) **synchronization** (2–4 seconds apart), and iii) **weak synchronization** (4–6 seconds apart). These three periods are displayed in Figure 1b as the vertical axis.

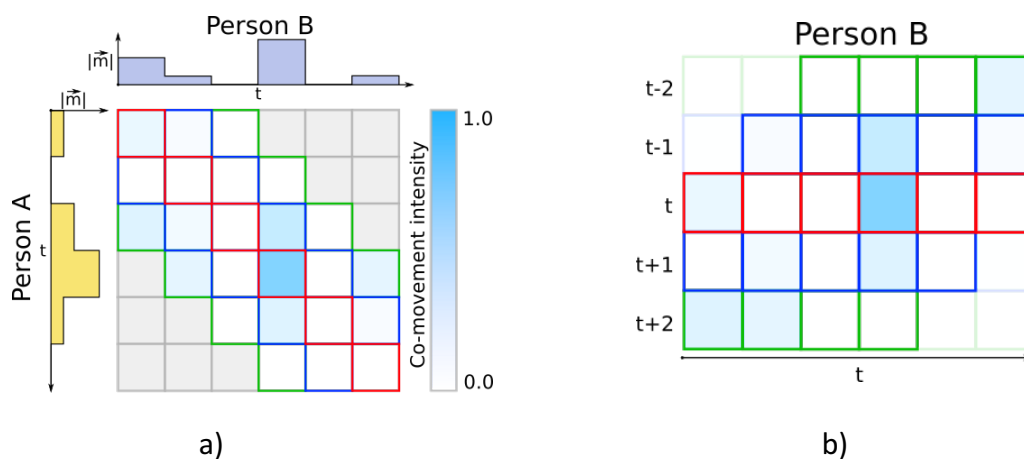a)   b)

Figure 1: Synchronized movement. a) Co-movement matrix of Person A and Person B over a period of 12 seconds (6 time steps). Perfect synchronization is represented by the diagonal of the matrix, marked with red squares. <±4 second synchronization is represented with blue cells and weak synchronization (<±6 seconds) is marked with green cells. Periods too far apart to be considered are grayed-out. b) Co-movement timeline, considered from the perspective of Person B. The figure shows the same values as the co-movement matrix, aligned on the diagonal cells of the matrix (red squares). Transparent sections are not present in the example matrix.

The additional third period was introduced to take into account indirect synchronization — when the person is not reacting to the teacher's stimulus but is following the reactions of others, for which we added 2 seconds for the person to observe the reaction of others and then reproduce it. This is what we call the "sleepers' lag" — the idea that those mimicking attention instead of actually paying attention will have a delay (a "lag") in their actions.

Algorithmically, motion synchronization between two persons was calculated as matrix multiplication. Each person is represented with a time series of motion intensity values, sampled in 2-second steps. The

co-movement matrix is created by multiplying the two time series as the *Nx1* and *1xN* matrix (visualized in Figure 1a). *N* represents the number of samples collected for each person during the lecture.

Within the two time series, values with the same index represent the same time-period in the lecture. This means that *perfect synchronization* moments will be found on the diagonal of the co-movement matrix, coordinates *(t, t)*. To analyze *synchronization* instances (2–4 seconds apart), Person A, who moved before, will occur 1 time-step before, and the co-movement with Person B is located at coordinates *(t-1, t)*. Similarly, "weak synchronization" with Person A moving 4 seconds before Person B is shown at coordinates *(t-2, t)*. In cases of mutual visibility, reverse direction of influence (Person B moving before Person A) is also possible and shown at coordinates *(t+1, t)* and *(t+2, t)*.

The majority of the co-movement matrix represents synchronized movement instances too far apart to be relevant (bigger difference between coordinates represents bigger time delays between actions). For that reason, we focus on the diagonal and the two bands around it: ±2sec, ±4sec. From the perspective of Person B, we can densely represent synchronization moments with Person A as the timeline shown in Figure 1b.

Because the values in the co-movement matrix represent multiplication of motion intensities in the range (0.0–1.0), the value produced will be high only if both movements were of high intensity.

4 METHOD

Our setup and method for gathering data is novel in the classroom environment. We will describe the main technological points, cover the data-gathering methodology, and provide our current working sample.

a) b) c)

Figure 2: Motion detection and grouping. a) Individual motion vectors shown as purple arrows, b) motion vectors grouped over time into motion tracks that can be assigned to an individual, and c) marked student areas and centres of Gaussian probabilities, which model the probability of motion belonging to each student.

4.1   Motion Analysis

Analysis of motion is based on tracking feature points in the video (Bouguet, 1999). Our setup consists of three cameras used for coverage of the students (shown in Figure 3) and one observing the teacher. Initial steps of analysis — synchronization of video streams from all sources and annotating visible regions in which students reside during the lecture — are described in Raca and Dillenbourg (2013).

Our main challenges in the process of extracting a measurement of motion for further analysis were i) interpersonal occlusions, ii) perspective distortion, and iii) normalization of the amount of movement recorded from a single person into a comparable measurement between several persons.

i) Interpersonal occlusions are handled by taking several pre-processing steps before assigning the motion to a person. The main idea is that by grouping the motion vectors into motion tracks, we can more reliably assign the whole track to a single person, instead of taking each motion vector as an isolated measurement.

Steps of the process are illustrated in Figure 2. Raw motion vectors are shown in Figure 2a as purple arrows whose intensities add to the amount of motion of one person at one time instance. Motion vectors (v) are next grouped into tracks (T) which consist of "cloud" of motion vectors over several frames. The criterium for grouping is based on proximity, direction similarity, and intensity of the vectors. For visualization purposes, a set of cloud centres from several frames are connected into a track, shown in Figure 2b. Finally the entire track is assigned to the student of highest probability (g_f), defined by the formula below. Each student (g) has a Gaussian distribution centred on the position of his head (depicted in Figure 2c). The entire track is assessed over every centre (i.e., every student) and motion is assigned to the student with the highest probability.

$$g_f = \underset{g}{argmax} \sum_{\forall v \in T} p(v \mid g)$$

In cases where a student was occluded on more than 80% of tracked area, the movements were indistinguishable from the person in front of him/her. Depending on the quality of the measurements for the person in front, either one or both students were removed from further processing if they were below a set threshold.

Taking into account that our primary interest was motion between students, it is important to notice that this method was designed so that

- A motion occurring between two students would not be assigned to both students,

- Large motions spanning several tracked areas would be assigned to a single person, and not to a group of people.
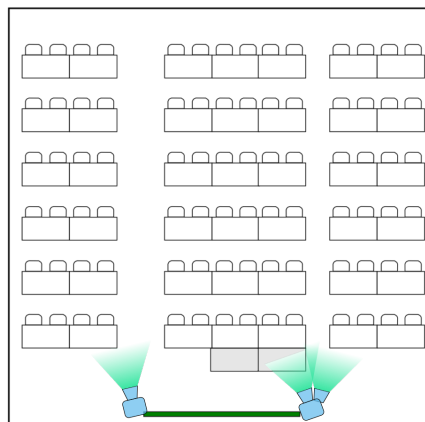
Figure 3: Arrangement of cameras for recording student actions.

ii) Perspective distortion: To compensate for the perspective effect, the number of tracking points remains constant over all annotated tracked areas. Our second precaution was to normalize the intensity of the motion vector by the diagonal of the student region. This ensures that the hand-motion of the student in the back row will be registered with the same intensity as the hand-motion of the student in the front row.

iii) Normalizing the amount of motion of a person has proven to be difficult. We based our normalization on two premises: i) the student is, on average, sitting still during the class; ii) the student has at least one full-body movement in the recorded footage (e.g., pose shift). To scale this to a range of 0–100% motion, we take the median value of movement intensity as the 5% motion (which corresponds to a small motion/sitting still being registered as 5% motion), and we verify that given this basic motion intensity the student reaches 100% motion at least once during the class. Motion that registers above the threshold of 100% is clipped to the maximum value. The final motion intensity over time can be visualized as shown in Figure 4b.

4.2   Experimental Procedure

We observed each lecture for the duration of 30 minutes. After a random interval (average duration 7 minutes) a tone signal was given that interrupted the lecture. At that time, students were asked to fill out a questionnaire sampling their activities and self-reported perception of the classroom. In addition to student samples, we hand-annotated class events that were products of teacher action or teacher–student interaction. Events were annotated into following categories: i) slide change, ii) slide animation, iii) question begin/end period, iv) answer begin/end period, and v) other events. Our questionnaire filling periods (typically lasting around 1 minute) were designated as "question answering" periods. Since they do not represent a normal part of a lecture, student activity in those periods was not taken into consideration in further data analysis. The events are shown as annotations in the top part of the timeline visualization in Figure 4b.
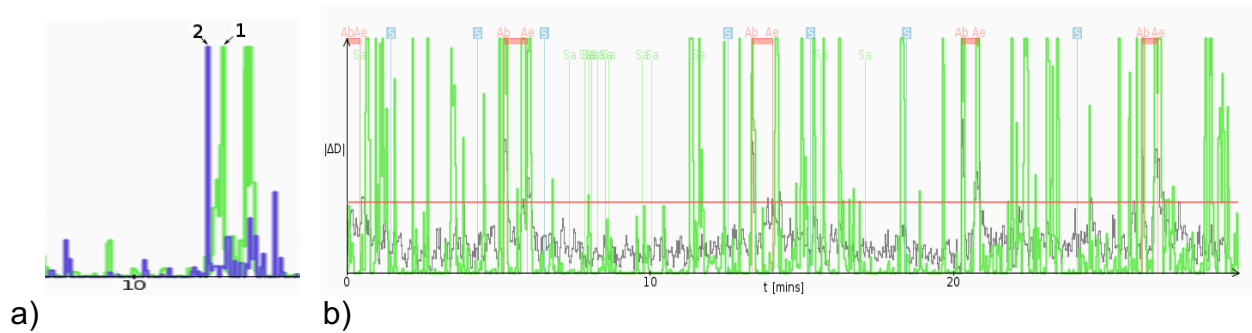
Figure 4: Motion intensity graphs. Horizontal axis represents the time and vertical axis 0–100% of relative motion of the person. a) Example of co-movement for two persons. Person 2 shifted her seating position (blue line), 2 seconds later, neighbouring Person 1 (marked in green) also started re-adjusting herself. b) Motion of a single person (dark green trace) overlaid on the average motion of the whole classroom (gray trace). The horizontal red line marks the 30% threshold that we used for movement analysis. Colour-coded labels on top indicate different events during the class, as described in Section 247.2. Annotations present here are: *Blue rectangles* — slide change; *Red periods* — question answering periods or questionnaire filling periods; *Green vertical lines* — slide animations. Questionnaires

By using a 10-point Likert scale, participants registered the following:

- their attention level
- their perception of the teacher (energetic/boring)
- their perception of the classroom attention (high/low)
- the importance of the material presented (important/irrelevant)

In addition to this, the questionnaire enumerated activities that the students did during the previous time-period:

- listening
- taking notes
- repeating key ideas
- thinking about other things
- interacting with people around you (if not scheduled by the classroom activity)
- using your laptop/phone

Students could check more than one activity.

4.3   Student sample

We base our results on analysis of two classes, described in Table 1. Both student groups were in the bachelor program of École polytechnique fédérale de Lausanne (EPFL). The teachers were two experienced

lecturers teaching social science (Class 1) and technical science (Class 2). The lectures were given at different times of the day — one in the morning, the other in late afternoon — and in different rooms.

Table 1: Basic information about analyzed classes.

Class	Size	Analyzed	Female ratio	Rows	Columns
1	38	29	36.8%	6	7
2	18	14	22.2%	4	5

Even though we initially consider both classes comparable, the small number of students in the second class rendered conclusions from that observation statistically invalid. We show the results found in Class 2 here to demonstrate the consistent trend in both cases.

4.4   Location and Surroundings

One of our main considerations when thinking about how the student perceives the lecture came from proxemic zones (Hall & Hall, 1969). Since the perception of the teacher changes significantly depending on how far the student is from the front, we decided against normalizing the space in the way it was done in Adams (1969), which would allow us to create one big sample by making the two classes comparable.

Emulating the proxemics concept in the classroom environment, we defined the three zones depicted:

- **Immediate neighbour** models "personal space." The person to the immediate left or right of the student, with whom the student shares desk- and leg-space. This is partially dictated by the dimensions of the desks, which in this case are made for two persons per desk.

- **Visible neighbourhood** represents the zone of two rows in front of the student 2 persons wide. This represents the "social zone" of proximal theory (which spans from 1.2m–3m). The zone models people who would be intentionally or unintentionally observed by the student following the material on the slides or looking towards the teacher.

- **Non-visible students** are those either too far to the side or behind the individual to be seen without intentional action.

5   OBSERVATIONS

5.1   Questionnaire Data

The collected questionnaire data was used primarily as the basis for further analysis of the collected video material. Nevertheless, we report the condensed findings to depict the general situation in the classrooms. A general note on the findings is that because of the small number of samples, we are reporting our findings with Kendall's correlation.

Reported levels of attention in both cases were high, with the mean located around seven. In the case of Class 1, μ=6.822, σ=2.344, and in case of Class 2 the normal distribution has the parameters μ=7.444, σ=1.100 (shown in Figure 7). This trend was also confirmed by our further studies using a larger sample of participants (194 participants), shown in Figure 7c, of μ=6.71, σ=1.456.

It is also interesting to observe that the attention reported was significantly correlated with the distance from the teacher (represented as the row in which the student was sitting). Responses shown in Figure 6 show the downward trend of correlation $r(192)=-0.29$ ($p<0.05$). This further confirms the observations of Daum (1972).

There is a significant correlation between the personal level of attention and the perceived level of attention of the entire class (Class 1: $\tau(38)=0.477$ ($p<0.05$); Class 2: $\tau(18)=0.413$ ($p<0.05$)). We considered this an interesting way of expressing dissatisfaction with personal or class performance as the student would mark a bigger difference between personal and classroom attention if there were a bigger dissatisfaction with the learning conditions. Classes were generally perceived by participants as exhibiting both high teacher-energy and high student attention.

Table 2: Parameters of perceived class quality.

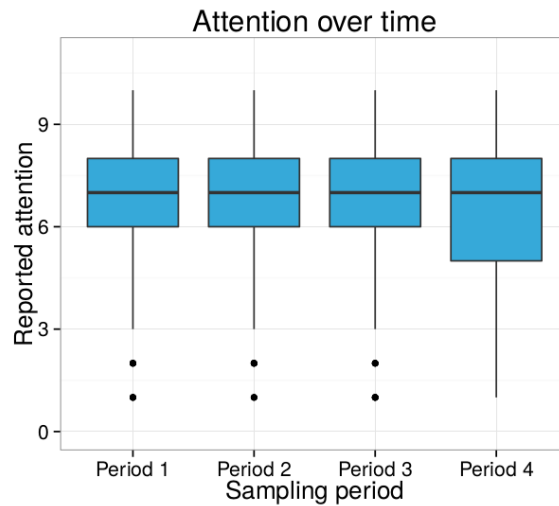| Class	Class attention (mean variance)	Teacher energy (mean, variance)
1	6.776 (3.711)	7.783 (1.866)
2	7.125 (3.266)	8.347 (1.920)

Figure 5: Attention over time at for 4 different moments during the class. Data captured in our extended study, sample size 194 participants.

We also studied the variation of attention levels over time in hopes of capturing the reported drop in concentration after 10 minutes (Wilson & Korn, 2007), but found no clear trend (see Figure 5).
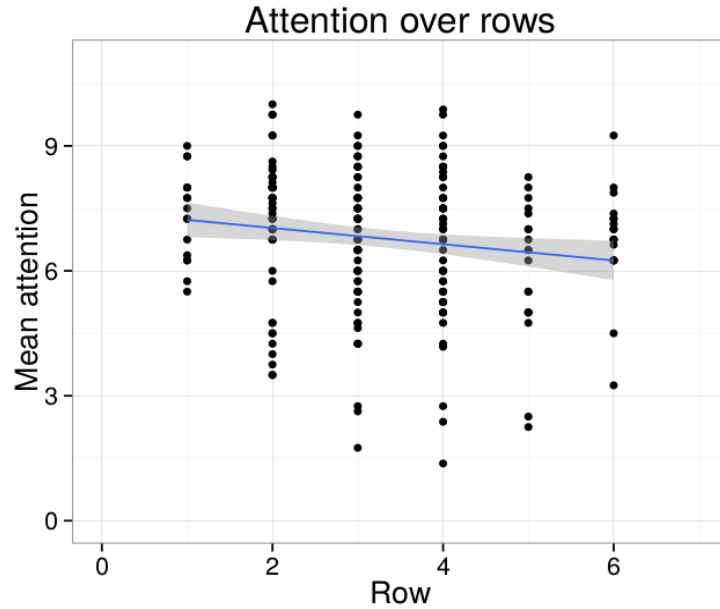

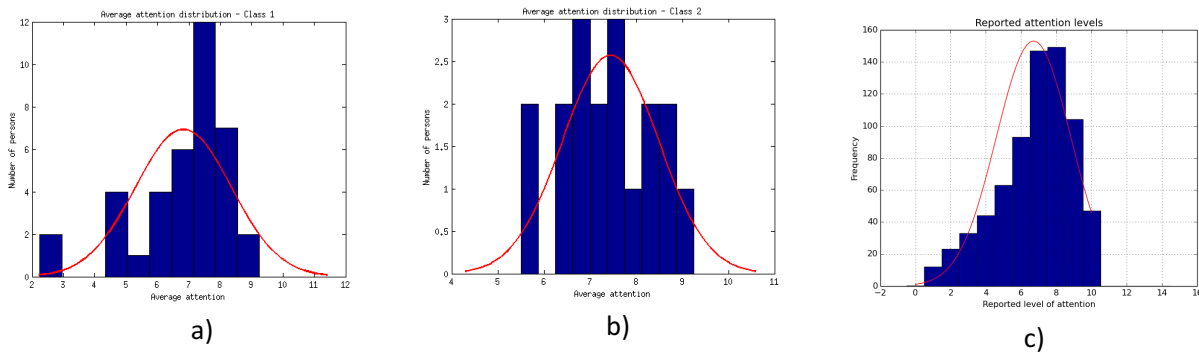**Figure 6: Mean attention reported over rows. Linear fit displays a slope -0.29, p<0.05.**

Figure 7: Average attention of students in both classes was subjectively perceived as high. *a*) Class 1 (μ=6.822, σ=2.344) and *b*) Class 2 (μ=7.444, σ=1.100) *c*) additional findings from our second study (194 participants) shows a cleaner Gaussian distribution (with the right-side tail cut-off) (μ=6.71, σ=1.456)
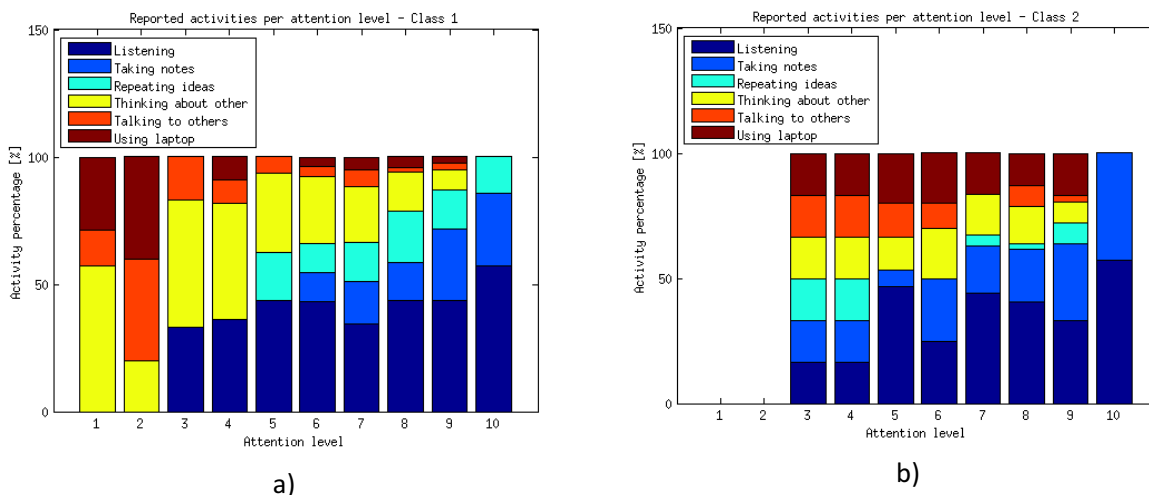
Figure 8: Percentage of activities per attention level in a) Class 1 and b) Class 2. Number of reported instances was normalized by the total number of instances on that attention level to produce the percentages.

Activities students reported (shown in Figure 8) show an expected tendency to report material-related activities (listening to lecture, taking notes, and repeating ideas) in higher attention levels. Off-task activities ("thinking about other things," "talking to others") were reported on all levels up to the maximum level of attention. Note that the students in Class 2 were using tablets as part of their regular studies to view the class material, which was not required for Class 1.

5.2   Motion Data

Synchronized movement is defined as body movement with more than 30% intensity from each of the two persons being compared (shown as the horizontal red line in Figure 4b). The 30% threshold was taken to separate minor body movements from motion likely to be noticed by others. We took into consideration the visibility of the two persons, meaning that in order for the movement of Person 1 to be considered as a stimulus, it must be visible to Person 2. Visibility reasoning was done based on the seating location of the two persons.

We compared the average number of synced movements between pairs sitting immediately next to each other and other pairs. We found that immediate neighbours had a higher probability of synchronized movement than a non-neighbouring pair (using a t-test ($p \leq 0.05$)), shown in Table 3.

Table 3: Average number of synchronized moments between immediate neighbours and other pairs.

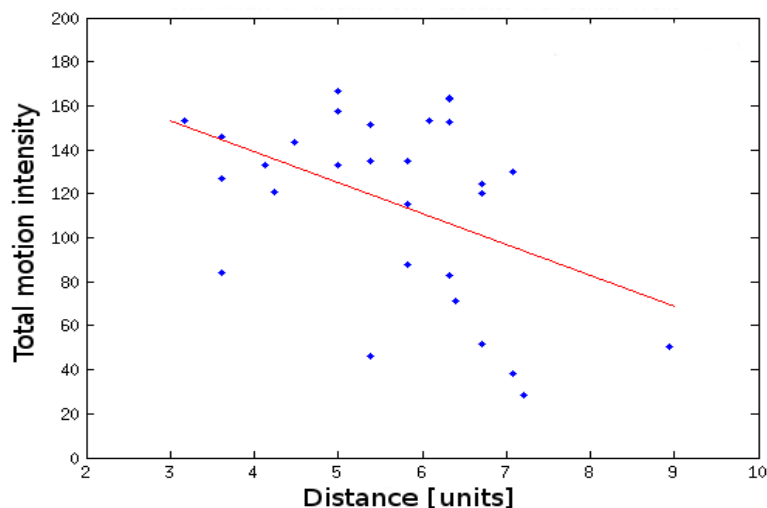| Class	Neighbouring pairs mean (variance)	Other pairs mean (variance)
1	76.54 (32.47)	54.43 (15.64)
2	63.33 (24.33)	44.88 (18.42)

Figure 9: Correlation between distance from teacher and motion intensity in Class 1; Kendall correlation t(38)=-0.284 (p=0.03)

We analyzed but found no significant difference in the number of synchronized movements between the pair from a visible neighbourhood and the non-visible students.

To compare the motion metrics with the previous findings of Adams (1969) on student activity, we also tested the influence of teacher proximity to the movement of the students. The further away students are from the front-centre of the classroom (the point closest to the teacher in both cases, represented as distance *d* in Figure 5) the less active they are (Kendall correlation is τ(38)=-0.284 (p=0.03) for Class 1; and τ(18)=-0.172 (p=0.45) for Class 2). Analyzing the samples, we have seen the same trend in both cases, even though the correlation was insignificant for the second classroom. Figure 10 shows the correlation for Class 1.
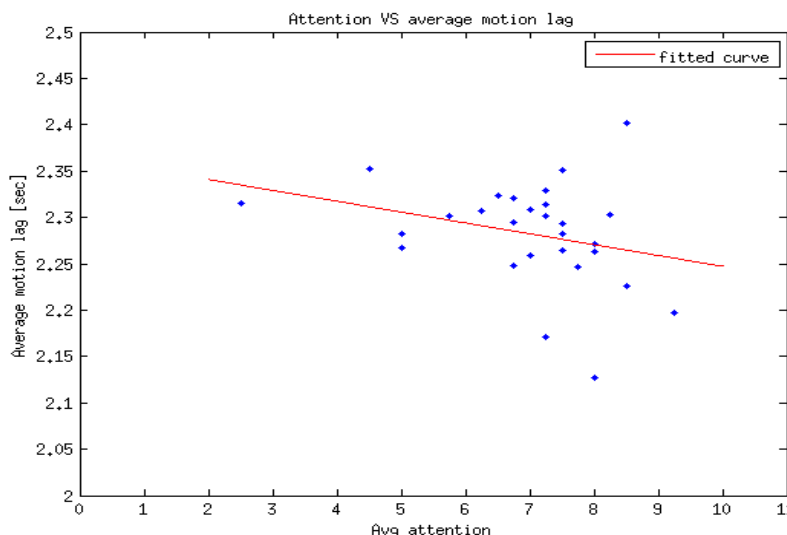

Figure 10: Average motion lag compared with the average level of attention in Class 1. Kendall correlation τ(29)=-0.259 (p=0.06).

Our third test was to find the correlation of the average reported level of attention to the reaction speed. The question was whether students with lower attention levels were more likely to lag behind other students in their visible field. The correlation found had the expected trend in the Kendall correlation ($\tau(29)=-0.259$ (p=0.06)) but was marginally insignificant. The result is shown in Figure 10. Class 2's correlation had a similar trend but was not statistically significant ($\tau(18)=-0.222$ (p=0.32)). The data thus suggests a phenomenon of "sleeper's lag," but the current sample is not conclusive. In addition, the difference in average speed of reaction is in sub-second intervals, which leads us to question if this would be noticeable to the teacher's eye without the technological enhancement of the classroom.

6   CONCLUSION

In this paper, we demonstrated our concept of measuring speed of reaction in the student population of the classroom. We gathered insight about the subjective perception of classroom attention with a questionnaire, which shows that students will project their level of attention onto others. Our first conclusion about synchronization of motion between immediate neighbours shows that two persons can affect each other just by sitting together without actual direct interaction.

We found a similarity with previous studies on the effect of teacher proximity on students (Adams, 1969; Daum, 1972) and found that students who are further away not only participate less, but also move less and report lower attention.

Finally, we proposed a new way of evaluating the overall attention of the classroom by comparing pairs of students and analyzing how synchronously they move. By comparing the motion results to the data gathered in the questionnaire, we showed a correlation between slower reaction time and lower levels of reported attention — the "sleepers' lag," but our data was not conclusive.

We have not yet touched on the subject of presenting the information to the teachers during the lecture, and we are planning to start a dialogue with the participating teachers to find the best representation for displaying the information during the lecture. Our next steps are to confirm the findings on a broader sample of students and continue to refine the technological methods. In addition to the "sleepers' lag" we would also like to explore further the phenomenon we call "distraction ripples" — assuming the transitivity of motion syncing, we would like to capture the spread of influence from one class-member to people around him/her. We are also interested in correlating how well these "ripples" spread in high-attention and low-attention groups of students in order to formulate a new metric of class attention.

In addition to motion, we aim to introduce additional cues into our reasoning about student attention and perception of the class, specifically gaze direction. The goal is to provide a holistic image of classroom life in order to find the most salient cues that can be unobtrusively collected. Our intention is that, in the end, the entire system would act as a training experience for novice teachers while also providing feedback to experienced teachers for continued professional development.

Stepping back from the trend of individual learning with massive online open courses (MOOCs), classrooms remain the dominant site of learning at all educational levels. Introducing technological

solutions to the classroom can potentially have a huge impact on the way students learn. By supplementing teacher observations with advanced measures, we hope to create a blend superior to current methods that exclude teachers, one that will be beneficial for students and teachers both.

ACKNOWLEDGEMENTS

REFERENCES

Adams, R. S. (1969). Location as a feature of instructional interaction. *Merrill-Palmer Quarterly of Behavior and Development*, *15*(4), 309–321.

Alavi, H. S. (2012). *Ambient awareness for the orchestration of collaborative problem solving* (Doctoral dissertation, École polytechnique fédérale de Lausanne, Lausanne). http://dx.doi.org/10.5075/epfl-thesis-5275

Allen, J. D. (1986). Classroom management: Students' perspectives, goals, and strategies. *American Educational Research Journal*, *23*(3), 437–459. http://dx.doi.org/10.3102/00028312023003437

Altman, I., & Lett, E. E. (1970). The ecology of interpersonal relationships: A classification system and conceptual model. In J. McGrath (Ed.), *Social and psychological factors in stress* (pp. 177–201). New York: Holt, Rinehart & Winston.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. *Proceedings of the 14th International Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (AIED '09), 17–24. Retrieved from http://centerforknowledgecommunication.com/newPubs/AIED SENSORS CameraReady.pdf

Bachour, K. (2010). *Augmenting face-to-face collaboration with low-resolution semi-ambient feedback* (Doctoral dissertation, École Polytechnique Fédérale de Lausanne, Lausanne). http://dx.doi.org/10.5075/epfl-thesis-4895

Bazzani, L., Cristani, M., Tosato, D., Farenzena, M., Paggetti, G., Menegaz, G., & Murino, V. (2013). Social interactions by visual focus of attention in a three-dimensional environment. *Expert Systems*, *30*(2), 115–127. http://dx.doi.org/10.1111/j.1468-0394.2012.00622.x

Bernstein, D. J. (2008). Peer review and evaluation of the intellectual work of teaching. *Change: The Magazine of Higher Learning*, *40*(2), 48–51.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, *26*(4), 473–498. http://dx.doi.org/10.3102/00028312026004473

Bouguet, J.-Y. (1999). Pyramidal implementation of the Affine Lucas Kanade feature tracker. *Intel Corporation*. Retrieved from http://robots.stanford.edu/cs223b04/algo_tracking.pdf

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education*, *6*(1), 9–20. http://dx.doi.org/10.1187/cbe.06-12-0205

Chen, D., & Vertegaal, R. (2004). Using mental load for managing interruptions in physiologically attentive

user interfaces. In *Proceedings of CHI EA'04, Extended Abstracts on Human Factors in Computing Systems,* 1513–1516. http://dx.doi.org/10.1145/985921.986103

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, *13*(1991), 127–149. http://dx.doi.org/10.1037/10096-006

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What Makes Great Teaching? Review of the Underpinning Research* (Research Report). Retrieved from the Sutton Trust website http://www.suttontrust.com/wp-content/uploads/2014/10/What-Makes-Great-Teaching-REPORT.pdf

Conigliaro, D., Setti, F., Bassetti, C., Ferrario, R., & Cristani, M. (2013). ATTENTO: ATTENTion Observed for automated spectator crowd analysis. In A. Salah, H. Hung, O. Aran, H. Gunes (Eds.), *Human Behavior Understanding: proceedings of the 4th International Workshop* (pp. 102–111).Berlin: Springer. http://dx.doi.org/10.1007/978-3-319-02714-2_9

Corcoran, R. P., & Tormey, R. (2012). *Developing emotionally competent teachers: Emotional intelligence and pre-service teacher education.* Oxford: Peter Lang/ERIC.

Daly, J. A., & Suite, A. (1981). Classroom seating choice and teacher perceptions of students. *The Journal of Experimental Educational*, *50*(2), 64–69. http://dx.doi.org/10.1080/00220973.1981.11011803

Daum, J. (1972). *Proxemics in the classroom: Speaker–subject distance and educational performance*. Paper presented at the annual meeting of Southeastern Psychological Association.

Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In Khine, M.S, I. M. Saleh (Eds.), *New science of learning: Cognition, Computers and Collaboration in Education* (pp. 525–552). Berlin: Springer. http://dx.doi.org/10.1007/978-1-4419-5716-0_26

Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., Bonnard, Q., … Kaplan, F. (2011). Classroom orchestration: The third circle of usability. In H. Spada, G. Stahl, N. Miyake, N. Law (Eds.), *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning* (CSCL 2011), (Vol.1, pp. 510–517). Lulu: ISLS.

Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. A., & Schroeder, U. (2013). Supporting action research with learning analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 220–229. http://dx.doi.org/10.1145/2460296.2460340

El Kaliouby, R., & Robinson, P. (2004). Mind reading machines: Automated inference of cognitive mental states from video. *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics* (SMC 2004)*, (Vol. 1, pp. 682–688). http://dx.doi.org/10.1109/ICSMC.2004.1398380

Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, *36*(2), 103–112. http://dx.doi.org/10.1207/S15326985EP3602_5

Ericsson, A., K. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, *15*(11), 988–994. http://dx.doi.org/10.1111/j.1553-2712.2008.00227.x

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. *Review of Educational Research*, *73*(3), 321–368. http://dx.doi.org/10.3102/00346543073003321

Forsyth, D. (2009). *Group dynamics (5th edition)*. Boston, MA: Cengage Learning.

Gatica-Perez, D. (2009). Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, *27*(12), 1775–1787. http://dx.doi.org/10.1016/j.imavis.2009.01.004

Hall, E. T., & Hall, E. T. (1969). *The hidden dimension* (reprinted 1990). New York: Anchor Books.

Hargreaves, A. (2000). Mixed emotions: Teachers' perceptions of their interactions with students. *Teaching and Teacher Education*, *16*(8), 811–826. http://dx.doi.org/10.1016/S0742-051X(00)00028-7

Hattie, J. (2013). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Heylighen, F. (2002). Complexity and information overload in society: Why increasing efficiency leads to decreasing control. *The Information Society*, *4*, 1–44.

Horvitz, E., Kadie, C., Paek, T., & Hovel, D. (2003). Models of attention in computing and communication: From principles to applications. *Communications of the ACM*, *46*(3), 52–59. http://dx.doi.org/10.1145/636772.636798

Howard, J. R., & Henney, A. L. (1998). Student participation and instructor gender in the mixed-age college classroom. *Journal of Higher Education*, *69*, 384–405.

Jermann, P., & Nüssli, M.-A. (2012). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. *Proceedings of the 2012 Conference on Computer Supported Cooperative Work* (CSCW '12), 1125–1134. http://dx.doi.org/10.1145/2145204.2145371

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Koneya, M. (1976). Location and interaction in row-and-column seating arrangements. *Environment and Behavior*, *8*(2), 265–282. http://dx.doi.org/10.1177/001391657682005

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, *37*(1), 153–184. http://dx.doi.org/10.3102/00028312037001153

Middendorf, J., & Kalish, A. (1996, February). The "change-up" in lectures. *National Teaching & Learning Forum*, *5*(3), 1–5. http://dx.doi.org/10.1002/ntlf.10026

Moore, M. G. (1989). Editorial: Three types of interaction. *The American Journal of Distance Education, 3(*2). http://dx.doi.org/10.1080/08923648909526659

Nüssli, M.-A. (2011). *Dual eye-tracking methods for the study of remote collaborative problem solving* (Doctoral dissertation, École polytechnique fédérale de Lausanne, Lausanne). http://dx.doi.org/10.5075/epfl-thesis-5232

Raca, M., & Dillenbourg, P. (2013). System for assessing classroom attention. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 265–269. http://dx.doi.org/10.1145/2460296.2460351

Range, B., Duncan, H., & Hvidston, D. (2013). How faculty supervise and mentor pre-service teachers: Implications for principal supervision of novice teachers. *Leadership Preparation, 8*(2) 43–58.

Rapp, D. N. (2006). The value of attention aware systems in educational settings. *Computers in Human Behavior*, *22*(4), 603–614. http://dx.doi.org/10.1016/j.chb.2005.12.004

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, *18*(5), 407–413. http://dx.doi.org/10.1111/j.1467-9280.2007.01914.x

Rivera-Pelayo, V., Munk, J., Zacharias, V., & Braun, S. (2013). Live interest meter: Learning from quantified feedback in mass lectures. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 23–27. http://dx.doi.org/10.1145/2460296.2460302

Roda, C., & Thomas, J. (2006). Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior*, *22*(4), 557–587. http://dx.doi.org/10.1016/j.chb.2005.12.005

Rosengrant, D., Hearrington, D., Alvarado, K., Keeble, D., Rebello, N. S., Engelhardt, P. V., & Singh, C. (2012). Following student gaze patterns in physical science lectures. *AIP Conference Proceedings of 2011 Physics Education Research Conference* (Vol. 1413, p. 323). http://dx.doi.org/10.1063/1.3680060

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Shell, D. F., Brooks, D. W., Trainin, G., Wilson, K. M., Kauffman, D. F., & Herr, L. M. (2010). *The unified learning model*. Berlin: Springer.

Stiefelhagen, R., & Zhu, J. (2002). Head orientation and gaze direction in meetings. In Proceedings of *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (pp. 858–859). http://dx.doi.org/10.1145/506443.506634

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2008). *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration [BES].* Iterative New Zealand: Best Evidence Synthesis Program. Retrieved from http://www.oecd.org/edu/school/48727127.pdf

Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, *27*(12), 1743–1759. http://dx.doi.org/10.1016/j.imavis.2008.11.007

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schröder, M. (2012). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing*, *3*(1), 69–87. http://dx.doi.org/10.1109/T-AFFC.2011.27

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1999). *Unobtrusive measures* (Vol. 2). New York: Sage Publications.

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, *265*(3), 94–104. http://dx.doi.org/10.1145/329124.329126

Whitcomb, J., Borko, H., & Liston, D. (2008). Why teach? Part II. *Journal of Teacher Education, 59*(4), 267–272. http://dx.doi.org/10.1177/0022487108322199

Wilson, K., & Korn, J. H. (2007). Attention during lectures: Beyond ten minutes. *Teaching of Psychology*, *34*(2), 85–89. http://dx.doi.org/10.1177/009862830703400202

Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 94–101. http://dx.doi.org/10.1145/2460296.2460315