Towards Automated Competency Estimation for Math Education – An Eye Tracking and Emotion Analysis Study

Andrea Kohlhase[1], Michael Kohlhase[2]

[1]Information Management, University of Applied Sciences Neu-Ulm
[2]Computer Science, FAU Erlangen-Nürnberg

Abstract
Personalization relies on user models – representations of the user's competencies, preferences, and skills to adapt the system behavior to optimize interaction. But the anticipated gain in productivity is offset by the effort involved in collecting and maintaining said user model. This is particularly pronounced in systems like ALeA (Adaptive Learning Assistant, https://courses.voll-ki.fau.de/), where the learner models contain competency estimations for thousands of concepts among multiple dimensions – here Bloom's learning levels.

In this paper we present an exploratory study design that tries to determine whether close visual observation of learners can be used to elicit competency data automatically – a task human educators perform routinely when teaching small groups of learners and adaptive learning systems should be equipped to mimic – with the help of this study.

Keywords
adaptive learning systems, competence monitoring, eye tracking, emotion tracking, Bloom taxonomy.

1. Introduction

Personalization relies on **user models** – representations of a user's preferences, skills and even competencies to adapt the system behavior to optimize interaction. But the anticipated gain in productivity is offset by the effort involved collecting and maintaining said user models. This is particularly pronounced in systems like ALeA (Adaptive Learning Assistant), where the user models are learner models and focus on competency estimations for thousands of concepts among multiple dimensions – here Bloom's revised learning levels [AK09].

In a small classroom, good educators would be able to determine competency by mere visual inspection, so we deem the automation possible just being based on collecting and understanding visual data. In this paper we present an exploratory study design that tries to enable our understanding of visual and/or emotional patterns while learning.

MathUI 2024
andrea.kohlhase@hnu.de (A. Kohlhase); michael.kohlhase@fau.de (M. Kohlhase)
https://kwarc.info/kohlhase (M. Kohlhase)
0000-0001-5384-6702 (A. Kohlhase); 0000-0002-9859-6337 (M. Kohlhase)

2. Preliminaries

Before we delve into the study and preliminary results, let us set the context:

2.1. ALeA, Symbols, and Learning Levels

ALeA [Ber+23; ALeA] is an interactive and learner-adaptive learning assistant system that generates personalized learner interventions based a **learner model**: a per-learner function that assigns Bloom competency vectors to symbols from the domain model. In ALeA, a **symbol** is a object, property of, or function/relation between objects. A typical 90 min lecture in Math, CS, or physics introduces 10-50 of them. In the sources of the ALeA learning materials, content words – i.e. the words that verbalize the concepts of the underlying domain – are annotated with the corresponding symbols.

A **Bloom competency vector** is a six-tuple of estimated mastery probabilities; one for each **learning level** in the revised Bloom taxonomy [AK09], which describe the cognitive processes by which thinkers encounter and work with knowledge via action verbs:

i) *remember*, by recognizing or recalling

ii) *understand*, by interpreting, exemplifying, classifying, summarizing, inferring, comparing, or explaining

iii) *apply*, by executing or implementing

iv) *analyze*, by differentiating, organizing, attributing

v) *evaluate*, by checking or critiquing

vi) *create*, by generating, planning, producing

Note that while Bloom's taxonomy thinks the levels above as a hierarchy, the Bloom competency vector allows for e.g. applying a concept without understanding, which is important in mathematics and computer science. Bloom's revised model tries to capture some of these phenomena by associating the levels with distinct **cognitive dimensions**: **factual**, **conceptual**, **procedural** and **metacognitive** that describe how knowledge is stored in our different kinds of memory and the corresponding retrieval mechanisms. Even though these give an important facet, we disregard them (for now) in the Bloom competency vectors.

ALeA uses a combination of the domain and learner models to make interactions with learning materials learner-adaptive. For instance at the end of a section in the course notes, we can collect all symbols that are introduced in it (via the symbol annotations), ask the learner model which have Bloom competency vectors below a configured threshold, and field remedial problems there. In contrast to a traditional (static) document where such problems are aimed at "the average reader", ALeA can adapt them to the actual reader (as far as the learner model is accurate). Similarly, before a section we can collect all symbols that are used but not introduced in the section and use them to determine whether the reader is prepared for the material – and again deploy remedial interventions to close any gaps. Of course the learner model is updated from the interactions with the remedial problems, leading to more accurate estimations for future adaptivity.

ALeA focuses on the first three levels of the Bloom taxonomy, since they are easier to diagnose and operationalize, and seem to be sufficient for automated personalization. In our experience the particulars of the competency model does not play a major role for personalization, as long as the domain model is fine-grained enough.

ALeA uses a particular operationalization of the Bloom competency vector based on its domain model: In the sources of the learning objects in ALeA, the parts of texts, diagrams, and formulae that correspond to a symbol are annotated with a pointer to that symbol. The **domain model** is a knowledge graph that records e.g. terminological dependency – what other symbols is a given symbol defined from? – which in turn can be used to generate learning interventions and explanations from. The learner model is a crucial component in the generation process: ALeA tries to target the result to the particular learner model by eliding content deemed sufficiently known/mastered or selecting exercises (understand/apply) or flash cards (remember) at the border of mastery.

In particular, the sources of **tasks** (e.g. exercises and exam problems) are annotated with **objectives** – what competencies does the exercise/problem determine: a pair (c, s) of a symbol s and a learning level c – and **prerequisites** – what does the learner have to master to have a chance to achieve the objectives (see [Loh+23] for details).

For a given component $\ell(s, u)_c = (c, p_c)$ of the Bloom competency vector $\ell(s, u)$ for learner u and symbol s, c is one of the 3/6 learning level and $p_c \in [0, 1]$ is the probability that u can solve a task with objective (c, s).

2.2. Eye Tracking

Eye tracking technology enables the close observation of eye movement (typically of an individual) together with the respective focus points. The focus is not moved continuously, but in saccades (jumps), at the end of which the person might or might not fixate, that is, look at for certain minimal threshold in terms of time. The locations, where the eyes rest, provide the information for the interpretation of **eye gazes**. In eye tracking experiments, it is typically assumed that the **eye-mind hypothesis** (see [JC80, p. 330], for math [SL19]) holds: whatever a person is fixating, she is not only perceiving, but even cognitively processing. In many studies this technology is therefore also used for measuring attention spans. In many studies the **Think-Aloud Protocol** is used (asking participants to report about their thoughts), to make it highly probable, that the eye-mind hypothesis is indeed valid.

The eye movement can be visualized by drawing lines between the fixation spots, where the size of a dot at each spot represents its length. These are called **gaze plots**. It lets the experimenter not only understand whether some stimulus was looked at, but also in which order the participant focused on presented objects. Another very intuitive visual analysis can be created in the form of **heat maps**, where typically red areas are the "hot" ones (looked at for long or frequently), green ones are the yet perceived objects and yellow the area in the middle. All other pixels of the stimulus are not overlayed with a colored pixel of some opaqueness. This way one can interpret at one glance whether wanted or unwanted areas were gazed at heavily.

2.3. Emotion Tracking

Emotions are arguably one of the main specialties of human beings. Robert Plutchik (e.g., [Plu82]) was one of the first to develop a classification scheme for them. The following nine classes of **emotions** are established in most emotion theories nowadays: **disgust**, **anger**,

contempt, **fear**, **sadness**, **joy**, **confusion**, **surprise**, and **sentimentality**. Emotions can be affected by each individual's **traits** with respect to an external stimulus as well as by their individual **moods**. The underlying **valence** covers the trait aspect, that is, by the object's intrinsic appeal/attractiveness or repulsion/averseness. Gasper et al. argue in [KG19] for including the **neutral** affective state supporting the mood aspect when considering emotions, where neutral concerns the feeling of indifference, the "presence of the neutral effect". Note that neutral is not just in between the positive and negative pole of valence, as they can co-appear. For example, a professor is asked by a student about learning resources for the next term, as he wants to start the process early on. The professor might feel neutral about whether to start now or then (mood), but might also feel honored by have been asked for personal advice (trait). Another factor for assessing an emotion is the human's current **engagement** (or arousal), which is a descriptor for her actual attention and involvement level.

Emotion theories vary in their view on the autonomy of emotions. Some argue that they are direct reactions to external stimuli, others that they are direct reactions to cognitive processing of external stimuli and some entertain mixed approaches. The same variety we can observe with respect to how they approach the possibility of manipulation of emotions. But it seems undisputed that **facial expressions** give relevant cues to a human's emotional state. Particularly, Paul Ekman researched facial expressions for their true representation of experienced emotions. For this he studied voluntary and involuntary facial expressions. The first are learned within the first few years of life (depending on social and cultural norms) and are habitual, non-conscious actions. Ekman notes that not all kinds of facial expressions correspond to emotions. Specifically, he (and others, see e.g. [WS02]) distinguishes

1. *conversational signals* which create context information for a conversation with other humans [Ekm86, p. 127],
2. *facial emblems* with an exact semantics (for instance, a one-eye closure wink),
3. *facial manipulators* that influence other parts of the face (e.g., lip sucking), and
4. *emotional expressions* which are involuntary signals carrying important information in social communication, indicating currently felt emotions.

Emotion tracking is based on observing facial emotional expressions. Some of these can be manipulated, but especially the micro-expressions, i.e., quarter-second occurrences of specific facial muscle switches, cannot. Ekman and Friesen [EF03] introduced the **Facial Action Coding System (FACS)** in 1978 and revised in 2002 to describe all visually distinguishable facial movements. Each such movement is decomposed into action units (AUs). Groups of these are interpreted in an emotion tracking system as emotions.

We call the result of combining eye tracking and facial expression data **emotional gaze data**. As mentioned above, we believe that educators use emotional gazes to refine their teaching to individuals, so we really want to study those emotional gazes.

3. The Study Design

The goal of the "Data-to-Bloom" study is to determine whether there are observable patterns in facial expressions that correspond to Bloom competency vectors when learners interact with particular learning objects.

For this goal it is insufficient to observe people reading learning materials: humans perceive and interpret facial expressions all the time, but we don't do so consciously, so we would miss out on a lot of data by mere human observation. Also, we could not directly see, which concept the observed person is looking at when modifying her facial expressions, so we couldn't get a grasp on emotional gaze data at all.

The combination of eye tracking and emotion technology, on the other hand, provides an optimal setting to research this, as we know exactly where a participant looks at any point in time and we can correlate that exactly with the facial expression at that point during an experiment. Assuming the eye-mind hypothesis, we can observe which concepts are being processed when which facial expressions appear, that is, we can observe emotional gazes. Unfortunately, it is not sensible to distract participants by asking them to apply the Think-Aloud Protocol method, if one is interested in potential time-dependencies between eye gazes and facial expressions, so the eye-mind hypothesis has to be taken as granted.

Both technologies provide us not only direct visual analyses for manually establishing hypotheses, but also gaze and AUs data to be used for confirming/refuting the hypotheses by other means.

3.1. Eliciting Learning Levels

We focus on eliciting the first three Bloom levels in Bloom competency vectors in this study. This poses the following specific challenges:

remembering To establish a kind of ground truth for this level we have to use concepts in a math context that most participants will probably remember. Recall that the components of a Bloom competence vector are probabilities, so we also need to elicit values for "not remembering".

understanding For a remembered learning concept we need to verify whether the participants can understand it. Note that we cannot simply create an exam question for that, as we would change the context of "studying learning material" to "testing understanding", which changes the underlying mood, trait, or emotion. In particular, we want to find a pattern for the depth of understanding while reading already internalized material on at least a remember level.

applying Here, we have to "make the participant think" and do the application. Therefore, we need to find a way of providing an application task without explicitly stating a task – again since we do not want to change the "studying learning material" context. Moreover, optimally we can guarantee that the participant's pre-knowledge is already on an understanding level.

So far, we haven't specified the concrete **target group** for this experiment – learning is not restricted to a certain population or culture. Note though, that by having established the above challenges, we need to switch from learning levels to knowledge levels. In particular, it would be far easier to confirm knowledge levels by restricting the target group to a math-educated population on a Bachelor level, so that we can assume abilities to reach up to Bloom's applying level.

To interpret the knowledge level in the emotional gaze data correctly, we need to verify the quality of the distinct knowledge levels for the concepts offered. As this cannot be answered in general, we set up a **post-test** (a test after the original experiment) in the study design, to establish self-assessed pre-knowledge-levels as well as the individual's potential mood and the concept's trait.

3.2. The Study Tasks

In the following, we describe the tasks we set our participants with respect to the Bloom levels we focused on, and explain, why we did it exactly that way.

Task 1: Not Remembering Here, the challenge consisted of finding a learning object, which is not known to (at least) most of our participants. If we were to select one from an advanced theory, then on the one hand, we would have to restrict our target group further to non-experts in that specific advanced theory, or on the other hand, we were to opt for an advanced theory from a far-away field like poetry. But the latter would potentially change the participant's affective states, so we decided against that option as well.

As a solution we invented a learning object, which sounded as if it could be an existing, but not-yet-known concept: The Cartesian Theorem. The theorem name itself would be enough as a trigger for a retrieval-from-memory process. Moreover, the name felt very familiar, so that participants would not directly quit the retrieval process.

To force each participant to reflect about this unknown concept without distraction, the stimulus consisted of only the string "The Cartesian Theorem" (see Figure 7). We prohibited an unwanted early escape by not offering the participant to advance to the next stimulus by herself - the system did it automatically.

We posed this task right at the beginning, to keep the expectancy level rather low, as they couldn't draw conclusions from prior experiences within the experiment.

Task 2: Remembering Pythagoras' theorem is a well known concept even in high school, so we started off with the presentation of the string "The Pythagorean Theorem" only (as with "The Cartesian Theorem") to directly observe difference in facial expressions for not-remembering and remembering. By eliciting the pre-knowledge level for each concept in the post-test we make sure that these levels hold.

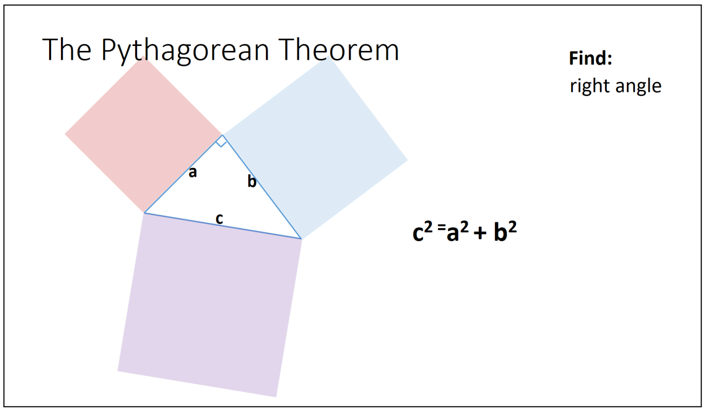


Figure 1: Remembering the Pythagorean Theorem

Then we intensified the participant's recall by adding more and more details and provide a finding task (see e.g. Figure 1) including uncommon names like "catheti". All the advances are done automatically by the system to have enough time for observing the emotional gazes.

Note that the equation itself is not properly formatted to verify the hypothesis that already known objects are only glanced at and not any longer studied in detail.

Task 3: Understanding After having reminded the participants at the content of Pythagoras' Theorem, we can now drive the experiment towards testing for the understanding level.

The next group of stimuli centers around the fact that Pythagorean triples (natural numbers a, b, and c, such that $a^2 + b^2 = c^2$, e.g. 3, 4, and 5) can be used for measuring/constructing right angles – a traditional application of the theorem of Pythagoras.

A famous artist, Friedensreich Hundertwasser, suggested that right-angle-architecture *lessens* creativity.

Now, do YOU live/work in a right-angled room?
To find this out you only have to have
- a measuring tape and
- to remember the Pythagorean Theorem!

Figure 2: Motivating a Use Case for Pythagoras

The understanding level is characterized e.g. by being able to explain or interpret an explanation of an exemplary application. Even though one might be tempted to think of this as Bloom's applying level, it is just a reproduction and not yet a proof of being able to apply a concept by herself.

The use case is motivated by framing non-right angles in a room as a potential personal advantage of using the concept. In Figure 2 there is the stimulus for announcing the usage of the Pythagorean Theorem. Now, the stimulus for testing the understanding level itself consists of the presentation of its solution in Figure 3.

Basically, we see two opportunities for showing understanding:

1. Scanning the bulleted list with the textual procedural description of the application of Pythagoras, and
2. Checking the validity of the assumption for applying the Pythagorean Theorem and the equivalence of the explicitly stated computation and the Pythagorean formula.

- Go to the corner of your room
- Measure and mark 30cm from the corner point on one wall
- and 40cm on the other
- Measure the distance between those points.

According to Pythagoras, if the angle is a right angle, then the square of this distance should be equal to $30^2 + 40^2 = 900 + 1600 = 2500$, that is, the distance should be $\sqrt{2500} = 50$cm.

Figure 3: Understanding the Solution

Task 4: Applying As we have shown how to apply the Pythagorean Theorem to verify a right angle, we can assume that participants are able to apply it as a thought exercise for themselves in a new context.

Thus, the next task is a historical variant of application of Pythagorean triples above that many people are not aware of: In ancient Egypt surveyors use the idea and implemented it with a rope with equidistant knots (see Figure 4); if it was stretched according to a Pythagorean triple of segments a surprisingly accurate right angle ensues (the reason why the Pyramids are so exact).

We expect an aha-effect when reading the bulleted list as it is a conclusion by analogy to the measuring tape problem - for some it might come earlier, for some only explicitly at the end. Note that this is still a rather lightweight application challenge, as it is sufficient if a participant thinks she understood how she would do it. This can be felt even without real understanding how it works.

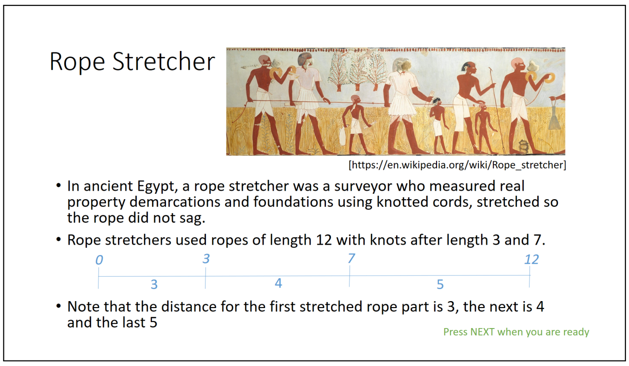


Figure 4: Applying: Conclusion by Analogy

The next stimulus tests for a deeper Bloom applying level: we frame it as a multiple choice problem (see Figure 5) to select for correct procedures to use a knotted rope and want to see how the emotional gaze might change. We leave the simple "studying learning material" as we don't know how to create this "thinking-the-application-through" situation otherwise. To confirm that this is harder than just the aha-effect, we ask in the post-test for self-assessment for these stimuli.

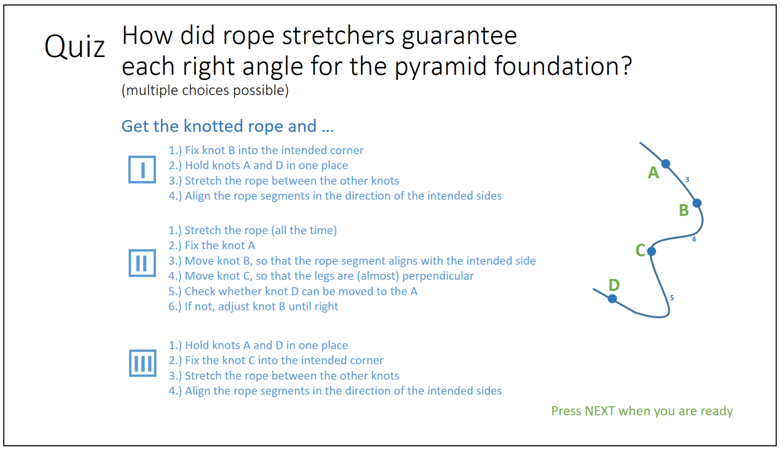
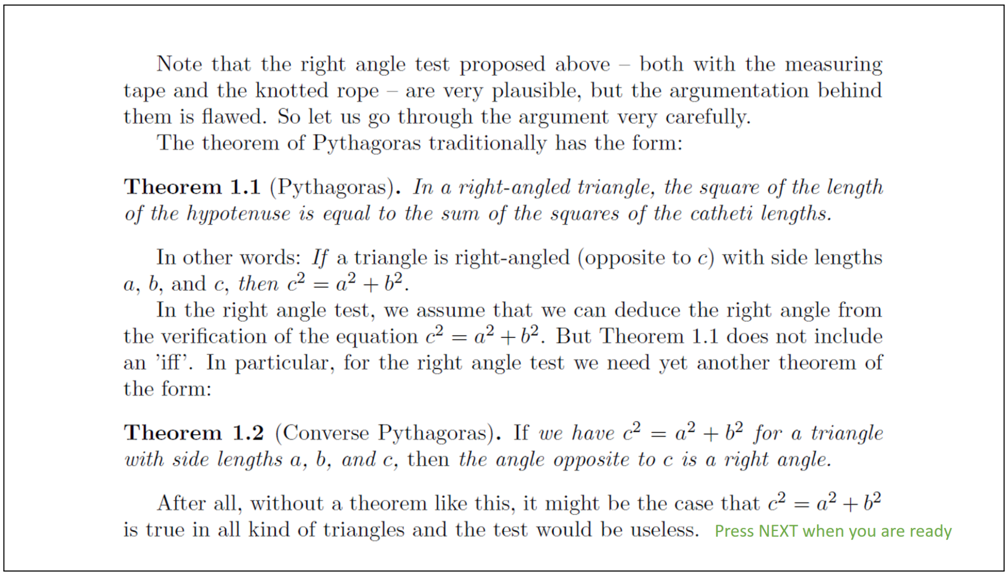


Figure 5: Applying Rope Stretching with Understanding

Task 5: Studying Learning Material In this task we pursue a wrinkle in the applications of Pythagorean triples that goes surprisingly un-noticed in high school education: Determining/constructing a right angle cannot be justified by the version of Pythagoras' theorem that is commonly taught: *If triangle* (A, B, C) *has a right angle at* C*, then* $|AB|^2 + |AC|^2 = |BC|^2$; we need the "only-if form" instead: *A triangle* (A, B, C) *has a right angle at* C*, if* $|AB|^2 + |AC|^2 = |BC|^2$ which is a simple consequence of the law of cosines.

In this task we change from the relatively lightweight formulation to a more mathematical exposition including explicitly given formulae and marked up/numbered theorems (cf. Figure 6), that is, we change to a real learning material style in a math context.

Here, we want to test our found and confirmed hypotheses in a more real world scenario.



Note that the right angle test proposed above – both with the measuring tape and the knotted rope – are very plausible, but the argumentation behind them is flawed. So let us go through the argument very carefully.

The theorem of Pythagoras traditionally has the form:

Theorem 1.1 (Pythagoras)**.** *In a right-angled triangle, the square of the length of the hypotenuse is equal to the sum of the squares of the catheti lengths.*

In other words: *If* a triangle is right-angled (opposite to c) with side lengths a, b, and c, *then* $c^2 = a^2 + b^2$.

In the right angle test, we assume that we can deduce the right angle from the verification of the equation $c^2 = a^2 + b^2$. But Theorem 1.1 does not include an 'iff'. In particular, for the right angle test we need yet another theorem of the form:

Theorem 1.2 (Converse Pythagoras)**.** If *we have* $c^2 = a^2 + b^2$ *for a triangle with side lengths* a*,* b*, and* c*,* then *the angle opposite to* c *is a right angle.*

After all, without a theorem like this, it might be the case that $c^2 = a^2 + b^2$ is true in all kind of triangles and the test would be useless. Press NEXT when you are ready

Figure 6: A Page in some Learning Material about Pythagoras

Task 6 (Experimental): Analyzing and Evaluating To already gain insights into the next Bloom levels we finish the experiment with yet another relatively unknown supposed application of the Pythagorean Theorem, where participants need to analyze the information and evaluate its correctness for themselves. We try to guide them by requiring them to evaluate a couple of statements. The guiding is arguably hidden by some non-standard phrasings, which make them harder to understand.

Concretely, the last task based on a news story where a journalist claims that the Pythagorean triples method (concretely with the triple 5, 12, 13) explored in the previous tasks was used before Pythagoras was even born.

3.3. Set-Up of the Study

The experiment was planned in a mobile setting, that is, like in a field experiment where participants conducted the test while being in a work location, which was due to pragmatic reasons.

We used a Tobii Pro Nano eye tracker with a sampling rate of 60 Hz[1] together with specific modules of the iMotions 10 Human-Factors Research software (10.0.35812.0), particularly the Core-, the Screen-Based Eye Tracking- and the Facial Expression Analysis Affectiva Affdex modules[2].

So far, we have confronted 50+ participants with 19 visual stimuli and 2 multiple choice tasks. All participants had already finished a Bachelors degree in either Mathematics or Computer Science. Many of them had even higher degrees. Accordingly, the age distribution is rather

[1] see details e.g. at www.srlabs.it/en/scientific-research/hardware-products/tobii-pro-nano/

[2] see e.g. imotions.com/products/imotions-lab/modules/

broad. 28 of the participants had an Indian cultural background, the others a Western one (distributed internationally).

4. Envisioned Analysis

To give the reader a flavor of how the analysis will proceed and what kind of results can be expected, we decided to present an exemplary preliminary hypothesis for Task 1 (see Section 3.2). To get a viable temporal pattern for the “not remember” situation that generalizes well the 12 affective states must be reduced to the relevant ones. Note that relevant does not mean, that the respective affective states have a high peak level. The pattern consists of a subset of states together with their observed peak level.

We studied 5 participants’ emotional gazes intensely and have manually identified four events of interest:

Peak1 The emotional snapshot directly after perceiving the concept name

Peak2 After realizing that there is no more to see and nothing to act on on the stimulus, so reflecting about the concept again (gaze is at the end of concept name)

Retrieval Effort Directly after Peak2 and taking up the serious effort to retrieve the content with respect to the concept name from memory

NotRememberState Concluding failed retrieval effort

The 5 participants were chosen randomly and we restricted ourselves to just five based on Nielsen’s "Discount Usability" principles, which advertise that observational patterns elicited by just 5 participants already cover 80% of the most relevant results ([Nie00]).

The left image in Figure 7 shows the emotional gaze of a typical participant for the stimulus "The Cartesian Theorem" right after she had fixated on the string "Cartesian" and glancing already at the name-finishing string "Theorem", that is at Peak1. Note that the vertical (red) line indicates at which point in time the eye gaze can be observed as a (yellowish) dot on the stimulus, which size depends on the length of the fixation. It is quite obvious, that here the emotions change for the first time, particularly Anger, Disgust, Fear, and Confusion high and Sadness on a middle amplitude. This could already indicate a specific subset of emotions to observe for the not-remembering level.

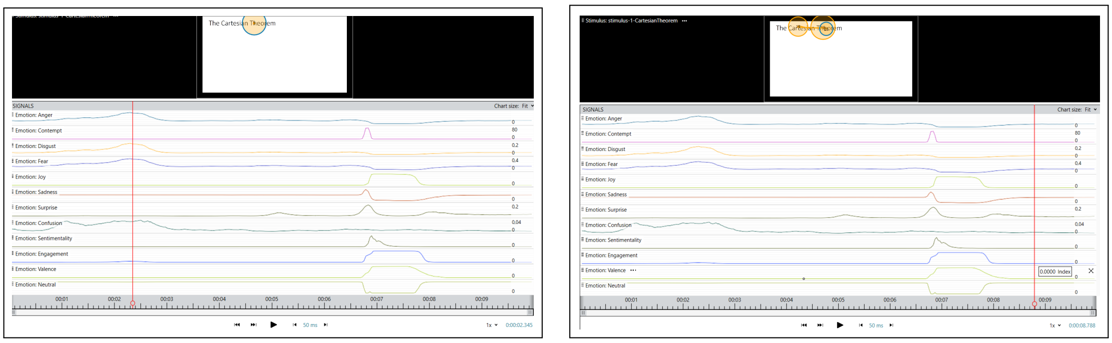


Figure 7: The Combination of Eye Gaze and Emotion Analysis

This specific stimulus is the first in the experiment and so we can expect a certain "gettingused-to-the-task" confusion. Therefore, this very first change of emotions might or might not yet belong to a pattern. But note the eye gaze in the right image of Figure 7 at the fourth event of interest: the same affective states are involved as at Peak1 (except for Confusion), but the peak level seems to be halved. This leads to a first hypothesis: *"Bloom's Not-Remembering level is characterized by a medium amplitude of Anger, Disgust, Fear, and Sadness"*.

These manually observed patterns will be feedbacked to our collaborators (Prof. Pradipta Biswas and his group at the Indian Institute of Science; see https://cambum.net/PB/), who will then use Machine Learning algorithms to confirm or refute the phenomenon based on significance over all participants' data for this stimulus.

5. Envisioned Application: Monitoring Competency in ALeA

If the qualitative patterns discussed in the last section can bear out quantitatively, we hope to use them to build a webcam-based monitoring subsystem that uses a combination of eye tracking and emotion analysis to interpret learner's facial expressions in terms of competency estimations for learner modeling in ALeA. This could replace tedious self-assessments in ALeA and thus significantly reduce the number of interaction to maintain learner models. Note that this also had a big effect on the learner model with respect to forgetting knowledge, which is typically not being self-assessed.

Concretely, learners would start the webcam of the device (typically a laptop or a mobile phone; both of which have one) they use for interacting with ALeA, which would send a stream of competency events (learner id, time interval, symbol, Bloom competency vector) to the ALeA learner model provider for integration. From our study, we can already see that on average there is at most one such update event per second, so this is a very low-volume reporting stream that is latency-uncritical, so we do not expect problems even over a low-bandwidth (mobile) connection.

The last years have seen the development of multiple light-weight, webcam-based eye trackers, some of them even in-browser and open source. Incidentally, IOS 18 due this year will feature eye tracking at the operating-system level for iPhones and iPads. So it does not seem unreasonable to develop the competency-monitoring system we envision over the next two years.

6. Conclusion and Future Work

We have presented an eye tracking and facial expressions study in math learning situations with the hope of finding patterns in the data streams that allow to infer competence estimations automatically. Initial qualitative analysis of the data from 50+ participants suggests that these patterns might be sufficient to build a non-invasive competency estimation facility based on webcams in mobile devices.

Note that the video data captured by the webcam in the monitoring application from Section 5 would stay entirely on the device and under control of the learner, and is thereby kept private. Only the competency data is transmitted to the ALeA learner model server – as would the competency data collected in more conventional ways. We feel that this instance of data minimization is an important desideratum in learning applications, even if it means that the processing has to be on the device.

Acknowledgments The work reported here was started while the authors were on sabbaticals hosted at the Indian Institute of Science (IISc) at Bengaluru, India. The intended application of competency monitoring in ALeA (cf. Section 5) has been inspired by systems from the research group of Prof. Pradipta Biswas and the study design has profited from discussions with this group and from IISc students as participants. The study has been completed while the authors were fellows of the Hausdorff Research Institute of Mathematics (Special Trimester "Prospects of Formalized Mathematics"), researchers there have also acted as study participants. The work was supported by the VoLL-KI project (see https://voll-ki.de) funded by the German Research/Education Ministry under grant 16DHBKI089.

References

[AK09] Lorin W. Anderson and David R. Krathwohl. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman, 2009.

[ALeA] *ALeA: Adaptive Learning Assistant*. URL: http://alea.education (visited on 07/16/2024).

[Ber+23] Marc Berges et al. "Learning Support Systems based on Mathematical Knowledge Managment". In: *Intelligent Computer Mathematics (CICM) 2023*. Ed. by Catherine Dubois and Manfred Kerber. LNAI. Springer, 2023. DOI: 978-3-031-42753-4.

[EF03] P. Ekman and W.V. Friesen. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Spectrum book Bd. 10. Malor Books, 2003. URL: https://books.google.de/books?id=TukNoJDgMTUC.

[Ekm86] P. Ekman. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. Berkley Books, 1986. URL: https://books.google.de/books?id=fBYRAQAAIAAJ.

[JC80] M.A. Just and P.A. Carpenter. "A theory of reading: From eye fixations to comprehension". In: *Psychological Review* 87.4 (1980), pp. 329–354.

[KG19] D. Hu K. Gasper LA Spencer. "Does Neutral Affect Exist? How Challenging Three Beliefs About Neutral Affect Can Advance Affective Research". In: *Front Psychol.* (2019). DOI: 10.3389/fpsyg.2019.02476.

[Loh+23] Dominic Lohr et al. "The Y-Model – Formalization of Computer-Science Tasks in the Context of Adaptive Learning Systems". In: *2023 IEEE German Education Conference (GeCon)*. IEEE, 2023. URL: https://url.mathhub.info/23GeCon.

[Nie00] Jakob Nielsen. "Why you only need to test with 5 users". In: (2000). URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/.

[Plu82] Robert Plutchik. "A psychoevolutionary theory of emotions". In: *Social Science Information* 21.4-5 (1982), pp. 529–553. DOI: 10.1177/053901882021004003. eprint: https://doi.org/10.1177/053901882021004003.

[SL19] Maike Schindler and Achim Lilienthal. "Domain-specific interpretation of eye tracking data: towards a refined use of the eye-mind hypothesis for the field of geometry". In: *Educational Studies in Mathematics* 101 (May 2019). DOI: 10.1007/s10649-019-9878-z.

[WS02] I. Wachsmuth and T. Sowa. *Gesture and Sign Languages in Human-Computer Interaction: International Gesture Workshop, GW 2001, London, UK, April 18-20, 2001. Revised Papers*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2002. URL: https://books.google.de/books?id=MCG-7-WCHu4C.