



# *An evaluation of an adaptive learning system based on multimodal affect recognition for learners with intellectual disabilities*

**Penelope J. Standen, David J. Brown, Mohammad Taheri, Maria J. Galvez Trigo, Helen Boulton, Andrew Burton, Madeline J. Hallewell, James G. Lathe, Nicholas Shopland, Maria A. Blanco Gonzalez, Gosia M. Kwiatkowska, Elena Milli, Stefano Cobello, Annaleda Mazzucato, Marco Traversi and Enrique Hortal**

*Penelope J. Standen, Emeritus Professor, Division of Rehabilitation and Ageing, University of Nottingham. Research interests in improving the quality of life for people with intellectual disabilities through psychosocial interventions and enabling technology. David J. Brown, Professor of Interactive Systems for Social Inclusion, Computer Science Team, Nottingham Trent University. Research interests in the use of enabling technologies to drive developments in special and inclusive education and in physical and mental health. Mohammad Taheri, Research Fellow. Interactive Systems Research Group, Computer Science Team, Nottingham Trent University. Research interests in affective computing and using multimodal data and machine learning to predict engagement in students with profound and multiple learning disabilities. Maria J. Galvez Trigo, Researcher, Faculty of Science, University of Nottingham. Research interests in the use of multimodal interactive systems and educational robots in the classroom to assist teachers and help children with Special Educational Needs and Disabilities. Helen Boulton, Emeritus Associate Professor in the Institute of Education, Nottingham Trent University. Research interests in Higher Education, Educational Theory and Educational Technology. Andrew Burton, Research Fellow. Interactive Systems Research Group, Computer Science Team, Nottingham Trent University. Research interests in the use of enabling technologies for students with learning disabilities and autism. Madeline J. Hallewell, Research Fellow, Faculty of Engineering, University of Nottingham. Research interests in stakeholder consultation including development of digital games using 3D audio. James G. Lathe, Research Fellow, Division of Rehabilitation, Ageing and Wellbeing, University of Nottingham. Research interests in methods for economic evaluation of clinical trials, particularly in the fields of mental health and wellbeing. Nicholas Shopland, Research Fellow. Interactive Systems Research Group, Computer Science Team, Nottingham Trent University. Research interests in the use of enabling technologies for students with learning disabilities and autism. Maria A. Blanco Gonzalez, Servicio de Equidad, Igualdad y Orientación Educativa. Consejería de Educación, Junta de Castilla y León. Monasterio Ntra.Sra.de Prado- Avda. Real Valladolid s/n- 47071 Valladolid. Research interests in technology provision for students with learning disabilities and autism. Gosia M. Kwiatkowska, Acting Director of RIX Research and Media Centre, University of East London. Research interests in enabling people with learning disabilities to participate in research to challenge their social exclusion, to have a voice and tell their stories. Elena Milli, Researcher, Polo Europeo Della Conoscenza, VIA PIO BRUGNOLI 36 Fumane, Italy. Research interests in supporting students with learning disabilities. Stefano Cobello, Coordinator, Polo Europeo Della Conoscenza, VIA PIO BRUGNOLI 36 Fumane, Italy. Research interests in intercultural education. Annaleda Mazzucato, PhD candidate Mind, Gender and Languages, University of Naples Federico II; Researcher and Project Manager, Fondazione Mondo Digitale, Via del Quadraro 102, Roma, Italy; Research interests in Social and Digital Inclusion, Technology Enhanced Learning, Human Mind and Cognition. Marco Traversi, CEO of Project Ahead. Has extensive experience in social innovation addressing the needs of disadvantaged people, including migrants. Founded the Italian Innovation Network. Enrique Hortal, Assistant professor, Department of Data Science and Knowledge Engineering Maastricht University, the Netherlands. Research interests in machine learning, human-machine interaction and affective computing. Address for correspondence: Penelope J. Standen, Division of Rehabilitation and Ageing, University of Nottingham, Nottingham, UK. Email: P.Standen@nottingham.ac.uk*

**Abstract**

Artificial intelligence tools for education (AIEd) have been used to automate the provision of learning support to mainstream learners. One of the most innovative approaches in this field is the use of data and machine learning for the detection of a student's affective state, to move them out of negative states that inhibit learning, into positive states such as engagement. In spite of their obvious potential to provide the personalisation that would give extra support for learners with intellectual disabilities, little work on AIEd systems that utilise affect recognition currently addresses this group. Our system used multimodal sensor data and machine learning to first identify three affective states linked to learning (engagement, frustration, boredom) and second determine the presentation of learning content so that the learner is maintained in an optimal affective state and rate of learning is maximised. To evaluate this adaptive learning system, 67 participants aged between 6 and 18 years acting as their own control took part in a series of sessions using the system. Sessions alternated between using the system with both affect detection and learning achievement to drive the selection of learning content (intervention) and using learning achievement alone (control) to drive the selection of learning content. Lack of boredom was the state with the strongest link to achievement, with both frustration and engagement positively related to achievement. There was significantly more engagement and less boredom in intervention than control sessions, but no significant difference in achievement. These results suggest that engagement does increase when activities are tailored to the personal needs and emotional state of the learner and that the system was promoting affective states that in turn promote learning. However, longer exposure is necessary to determine the effect on learning.

**Introduction**

The term intellectual disabilities (ID) has become the internationally recognised replacement for terms such as “learning disabilities” and “mental retardation” (Schalock *et al.*, 2010). There are three criteria to be met before a diagnosis of ID can be made: an intellectual impairment (IQ below 70), significant difficulty with daily living skills and onset before the age of 18 (American Psychiatric Association [APA], 2013). It can be sub divided into four levels: mild, moderate, severe and profound. In theory these levels are defined by IQ but in practice not everybody receives a formal testing of their IQ and it is more useful to think of people differing in the level of support they require, a categorisation proposed by the American Association on Intellectual and Developmental Disabilities (Luckasson *et al.*, 2002). One of the areas in which they need support is education, yet, students with ID are not always receiving appropriate, accessible and meaningful opportunities to learn (Taub, McCord, & Ryndak, 2017).

Artificial intelligence tools for education (AIEd) may be one strategy to provide the personalisation required for students with ID who experience such a wide range of learning needs and do not necessarily learn or develop in a linear or hierarchical way (Colley, 2013). AIEd is positioned to address the limitations of “one-size-fits-all” learning, with inflexible learning pathways (Nesta, 2019), whilst retaining the benefits of learning with a class cohort and those of personalised instruction. The use of multimodal data, for instance in affect recognition, leads to more accurate machine learning classification models to augment decision-making processes (Cukurova, Kent & Luckin, 2019). Given the role of affect in learning (Baker, D’Mello, Rodrigo & Graesser, 2010; Kort, Reilly, & Picard, 2001), one of the most innovative applications of AIEd is in this detection

### Practitioner Notes

What is already known about this topic

- Students with intellectual disabilities are not always receiving appropriate, accessible and meaningful opportunities to learn.
- AIED may allow the degree of flexibility and personalisation for students who are disadvantaged in the traditional classroom environment.
- The inclusion of affect recognition to optimise support for the learner or to personalise the presentation of learning material, has shown some positive results with higher education students.

What this paper adds

- This is the first study to develop and evaluate an adaptive learning system based on automatically detected affect for learners with intellectual disabilities who are working at a level far below their mainstream peers: a group rarely catered for in terms of AIED.
- Lack of boredom had the strongest link to learning achievement, with both frustration and engagement positively related to achievement.
- When compared with sessions using the adaptive learning system without affect sensing driving the presentation of learning material, learners were more engaged and less bored but there was no significant difference in achievement.

Implications for practice and/or policy

- Although future work is necessary to refine the machine learning method and develop a greater range of learning material, an adaptive learning system based on automatically detected affect that can support teachers of those with intellectual disabilities is one step nearer.
- Such a system that responded in real time to the learner's affective state allows teachers to decide on a more appropriate distribution of their close monitoring to ensure they can target their efforts so that all students will be supported to reach their full potential.

of a student's affective state, to "enhance learning by means of nudges that move students out of negative states such as boredom or frustration that inhibit learning into positive states such as engagement or enjoyment" (du Boulay, Poulouvasillis, Holmes, & Mavrikis, 2018, p. 23).

Recent reviews (eg, Yadegaridehkordi, Noor, Ayub, Affal, & Hussin, 2019) have highlighted an explosion of work in this area, identifying more than 20 different affective states (eg, frustration, confusion, boredom and engagement) from features such as facial expression, posture, skin conductance, heart-rate and brain signals. This information has been used in AIED systems in different ways.

One way is to provide real time information on engagement to teachers so that they can implement "just-in time" personalised interventions (Aslan *et al.*, 2019). Affective information has also been used to provide the assistance that a human tutor might provide via an agent (D'Mello & Graesser, 2012a). Compared to controls, this type of assistance was associated with significant learning gains in struggling college students. Thompson and McGill (2017) used the affective state of university students to drive the actions of an agent providing guidance and support as would a human tutor. All measures of effectiveness were higher for those students receiving affective support compared with those using identical software with the affective components disabled, but only significantly higher for perceived and not actual learning. The *iTalk2Learn* intelligent

learning platform designed for 8–12 year old children who are learning fractions, detects affective state through speech analysis and uses it to determine the timing and type of feedback message. Compared with performance determined feedback, affect aware feedback reduced boredom and off-task behaviour (Grawemeyer *et al.*, 2017).

Information collected can also be used to drive the presentation of learning materials so that it adapts to the learner's current needs. Scheiter *et al.* (2019) developed a gaze-contingent adaptive system that analysed learners' eye movements. When poor information processing was detected, presentation of the materials was altered to trigger a more adequate type of processing.

The case for applying AIED in those with ID may be even stronger, where schools are receiving more diverse students in their classrooms requiring diverse teaching. Approaches that address the real issue of teachers not having enough capacity to attend to each child's individual learning needs are called for, to ensure that all students are supported to develop their full potential. Online education programmes can provide a variety of multimedia and flexible scheduling (Rose & Blomeyer, 2007), which would allow individualised instruction to meet the specific needs of the most cognitively challenged learners. Learners can progress through learning material at their own pace, spend as long as is needed on concepts that have not been fully grasped but skip over those that have (Bertini & Kimani, 2003). Via the use of AIED technologies, an even greater degree of personalisation can be achieved, but development for learners with ID in this area is less prolific than for mainstream learners. Approaches to designing learning environments to meet the needs of learners with multiple disabilities have been suggested (Nganji & Brayshaw, 2017), by analysing the needs of the learner and then, matching specific learning resources to individuals through ontological modelling and adaptive personalisation using basic machine learning concepts. Evaluation of a learning environment in which children diagnosed with Autism Spectrum Conditions (ASC) engage in social interactions with an artificially intelligent (AI) virtual agent showed a significant increase in the proportion of children's responses to the human social partners when acting in support of these interactions (Porayska-Pomsta *et al.*, 2018). However, these do not consider the learners' emotional states when guiding the personalisation process. Research has shown that engagement increases when activities are tailored to personal needs and emotional states (Athanasiadis, Hortal, Koutsoukos, Lens, & Asteriadis, 2017). Effectiveness of the learning process has been directly correlated with learners' engagement in learning activities (Hamari *et al.*, 2016).

### **An adaptive learning system based on affect sensing (the MaTHiSiS system)**

In order to investigate the effect of a system which personalises learning activities based on learners' needs and their emotional states, the MaTHiSiS project (<http://mathisis-project.eu/>) adopted an innovative modelling strategy for learning experiences, multimodal affect recognition and an on-the-fly adaptation strategy (see Figure 1, Tsatsou *et al.*, 2018).

The MaTHiSiS adaptive learning system aimed to identify three affective states: engagement, boredom and frustration. The affective model used was based on the pedagogical framework for online tutoring outlined by Basawapatna, Repenning, Han Koh, and Nickerson (2013) and Taheri *et al.* (2018), which combined Csikszentmihalyi's Theory of Flow (Nakamura & Csikszentmihalyi, 2002) with Vygotsky's Zone of Proximal Development.

Students with intellectual disabilities especially those who also have autism, do not necessarily display the affect states relevant to learning in the same way as their nondisabled peers. This may be in terms of different facial expressions, posture and movements due to co-occurring neurological differences and different use of eye gaze in those with autistic tendencies. Self-annotation of recordings (eg, Chickerur & Joshi, 2015) was not appropriate in our study due to the ability of the participants. Therefore, rather than use a method established for main stream learners (eg,

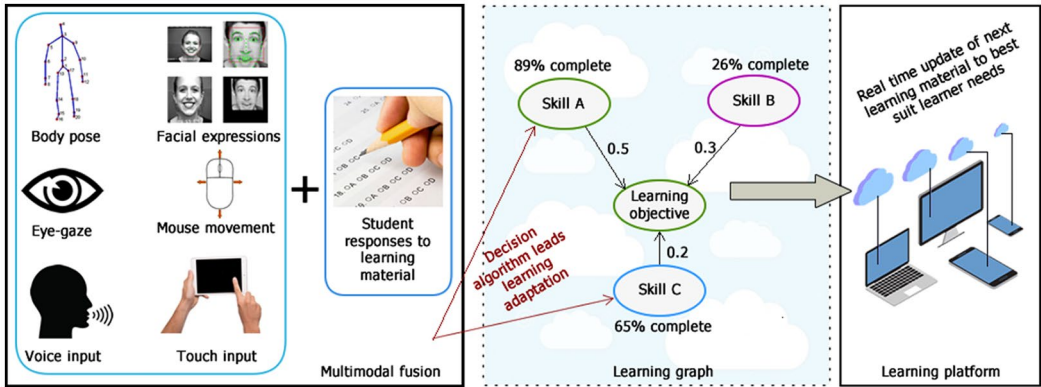


Figure 1: MaTHiSiS adaptive learning ecosystem

BROMP, Ocumpaugh, Baker, & Rodrigo, 2012), in order to determine the affective state of students with ID, teachers and trained researchers familiar with the students annotated recordings of them working on educational materials on a variety of platforms (desktop, mobile devices and educational robots).

These labels were used to train machine learning algorithms for a range of modalities including: facial expressions (pretrained and subsequently fine-tuned C3D Convolutional Neural Network (CNN) model to infer the emotions from face images enhanced via transfer learning using public data sets), eye gaze estimation (two-stream CNN using 3D gaze vectors), body pose (Speed Relation Preserving Slow Feature Analysis algorithm to extract features classified by a Support Vector Machine model (SVM)), voice input (classified through CNN architecture SVM using the labelled data set), gestures (mobile platform accelerometer and gyroscope data are computed to obtain 3D accelerations and jerk which are classified using a SVM model) and interaction with learning materials (information regarding the students (skill level, affective state, etc), their interactions with the platform including touch input and mouse movement (time per exercise, total time, number of trials, etc) and the platform itself (type of activity, level of difficulty)).

Each of the models for the various modalities was trained with the data gathered at the testing sites using a Data Acquisition Tool in the pre-experimental phase of the MaTHiSiS project. The performance of each were based on k-fold cross-validation for the majority of the modalities and leave-one-out cross-validation for the interaction parameters classifier. This was considered to be the most reliable approach since it would be agnostic to the different ways of using and handling the devices available to students (including tablets, robots and mobile devices) and the different models used (Ghaleb, Hortal, *et al.*, 2017a).

An equally weighted late multimodal fusion scheme using the predictions independently inferred by each modality was employed to give an overall understanding of the affective state of each learner as this approach has been shown to improve accuracy and reduce discrepancy in the recognition of affective states (D'Mello & Graesser, 2010). This late fusion scheme uses the modalities available after their potential rejection based on their usability according to the individual user profile. The final set of modalities are averaged using the probabilities per class and the emotion with a higher averaged probability is selected. Initially, a Genetic Algorithm solution was proposed to learn appropriate weights per modality and use case (Ghaleb, Popa, *et al.*, 2017b). However, due to the limited amount of data available to properly train such a complex architecture and the reduced impact on the final accuracy, the simpler solution mentioned previously was

applied. This approach reduced the training process and did not have an adverse impact on the fusion performance.

Adaptation of the learning process depends on the affective state of the learner and the use of “learning graphs.” Once the learner leaves a state of flow, the system automatically adjusts the challenge level of the learning material. If their state tends to boredom, learning material challenge is increased to induce interest. If anxiety is detected, the challenge level is lowered to relieve difficulty. Also, in the MaTHiSiS system persistent states of boredom or frustration are met with intervention by design. Persistent frustration is met with decrease of challenge or change of learning material and persistent boredom is met with increase of challenge or change of learning material. This adaption process aims to maintain the learner in a state of flow (Taheri *et al.*, 2018).

Learning graphs contribute to the adaption of the learning process by adapting goal weights, so that they reflect both the contribution of the goals to the overall learning objective, leading to a smooth transition in knowledge acquisition (Tsatsou *et al.*, 2018). Through sensing immediate changes in affective state, immediate changes in the presentation of learning material reduce the probability that the learner will slip out of an engaged state and stay out of this optimal state for lengths of time that might require more drastic interventions or that might alienate the learner.

The current study set out to evaluate the effectiveness of the MaTHiSiS adaptive learning system as a means for maximising engagement and learning in school aged children with ID by addressing the following hypotheses:

1. The sensor data can be used to automatically identify different affective states that were associated with learning achievement.
2. The MaTHiSiS adaptive learning system has a positive effect on engagement and learning achievement.

## **Materials and methods**

### *Design*

A within subjects repeated measures design was adopted whereby each participant took part in intervention (A) and control (B) sessions. The intervention (A) was MaTHiSiS used as it was designed: with affect and achievement data driving the presentation of the learning material and (B) where the presentation of the learning material was based on achievement alone. The advantages of this design are:

- Each participant acted as their own control, thus, controlling for differences between very varied participants.
- It was flexible enough to fit in with teachers’ and learners’ requirements as session length and timing of sessions can vary to suit classroom and learners’ obligations.
- It reduced the order effect that comes from one condition always being first or second.
- It maximised the number of testing sessions to minimise effects of any unwanted variations such as time of day or specific learning material.

### *Ethics*

Ethics approval was received from the University of Nottingham’s Faculty of Medicine and Health Sciences Research Ethics Committee, B16122016.

Table 1: Table of characteristics of participants

		Total (N = 67)	ID (N = 23)	ID/ASC (N = 22)	ASC (N = 22)
Age at 2018 in years	Mean (SD)	10.1 (2.6)	9.2 (2.1)	11.6 (2.9)	9.6 (2.0)
Gender N (%)	Female	21 (31.3)	6 (26.1)	10 (45.5)	5 (22.7)
	Male	46 (68.7)	17 (73.9)	12 (54.5)	17 (77.3)
Level of intellectual disability	None	7 (10.4)	0	0	7 (31.8)
	Mild	18 (26.9)	12 (52.2)	1 (4.5)	5 (22.7)
	Moderate	24 (35.8)	8 (34.8)	8 (36.4)	8 (36.4)
	Severe	18 (26.9)	3 (13.0)	13 (59.1)	2 (9.1)

### Participants

Participants were recruited from schools at six different sites: Nottingham and London in the UK, Rome, Salerno and Fumane in Italy and Valladolid in Spain. The inclusion criteria were:

- Working at a level way below their peers having either ID or with autistic spectrum condition (ASC).
- Aged between 6 and 18 years.
- Nominated by teacher for being able to potentially benefit from using the MaTHiSiS system.
- Having parental or carer consent to participate.

67 participants took part in at least one control (B) session and their data were included in the evaluation.

Participants were allocated to one of three groups according to information on their school record: those with intellectual disabilities only (ID), those with intellectual disabilities with some autistic tendencies (ID/ASC) and those for whom a diagnosis of autism was the primary presenting feature (ASC) (Table 1).

### Intervention

Teachers selected learning material from a library to create their own *Learning activities* and *Learning graphs*: an online equivalent of a specific lesson in traditional learning environments, where several learning goals are defined and are expected to be acquired. To reach these goals, the learning experience is divided into several *Smart Learning Atoms* which are representations of small pieces of knowledge (Boulton, *et al.*, 2019). These reusable learning objects are self-contained learning components that are stored and accessed independently. They can be reassembled to create new courses or sequenced to form individual learning paths. This level of granularity enables a higher degree of personalisation as changes in affect detection and presentation of learning material can be implemented more immediately. The library was created by teachers at the participating schools to be suitable for the students taking part (no or limited, verbal abilities). This approach also had the advantage of allowing the sharing of learning material between the different countries.

A range of Learning Objectives were embedded in the Learning Graphs (Figure 1) at the testing sites. These included improving navigation, sequencing and vocabulary skills and improving awareness of cause and effect in the UK testing sites. At the Italian testing sites, tutors developed Learning Graphs aimed at improving attention, language, maths, vocabulary, navigation and social skills. Learning Graphs to improve social, navigation, sequencing, vocabulary and maths

skills were developed for use at the Spanish testing sites (Mazzucato & Traversi, 2017). Learning graphs were differentiated based on the level of support required by the students and the different situational requirements of each testing site (for example it was not possible to use the NAO robot platform agent at some sites).

### *Outcome measures*

For each session, affective state was calculated as the proportion of time in which the participant spent in frustration, engagement or boredom as calculated from the affective state recognition software. Using the late multimodal fusion approach presented earlier, modalities are first selected based on the “user profile.” The classification outcome probabilities reported from each modality is averaged per affect class (“frustration,” “engagement,” “flow”). These probabilities were expressed as a figure between 0 and 1 for each state, with 1 representing when the model detects full presence of that affect state. The affect class that has the highest probability outcome (averaged from each modality) is then selected as the main affect class for that segment of multimodal sensor data.

Learning was expressed by the achievement value calculated by the MaTHiSiS software: the overall competence score that the learner is achieving over the entire learning activities worked on to achieve a particular goal. This figure is calculated from the number of correct and incorrect answers and ranges from  $-1$  to  $1$ .

### *Procedure*

Teachers and the supporting research team were advised to involve each participant in 12 sessions, half of which would be intervention. To reduce the order effect, teachers were advised to alternate sessions between the two conditions in bouts of three, that is, AAA BBB AAA BBB, with half of the participants experiencing a reversed order of the conditions, that is, BBB AAA BBB AAA. Teachers were advised to end the session whenever they thought appropriate for the learner, but to avoid going over 20 minutes. This pattern was selected in order to maximise collection of robust data without proving too onerous for teachers and learners. However, due to technical problems and learner absences due to illness or behaviour, this pattern could not always be adhered to precisely but alternation between conditions was maintained. Participants worked through learning graphs considered relevant for them by their teachers. The choice of device on which they interacted with the system (laptop, tablet or NAO robot) was determined by their teacher.

The number of A sessions ranged from 1 to 13 (mean 5.3), with 91% of participants taking part in 3 or more A sessions. The number of B sessions ranged from 1 to 11 (mean 4.3), with 75% of participants taking part in 3 or more B sessions. Total time during which the participant was using the system either in A or B sessions ranged from 15 to 413 minutes (mean 113 minutes). About 84% of participants had a total duration of 60 minutes or above.

### *Analysis*

Data were extracted from the MongoDB database, one file for affective state and one for achievement. For each session, the system outputted values for performance several times a minute, for affective state slightly less often. For each session, relevant lines of data were identified using the information provided by each testing partner in their posttesting spreadsheets. All the values for each affective state and achievement were averaged so that for each session there was one value for probability of engagement, boredom, frustration and achievement. Achievement was converted to a positive value by adding 1 to the recorded figure. Statistical analysis was carried out using SPSS v25 and Stata 15SE.



In order to address the first hypothesis (sensor data were automatically identifying different affective states associated with learning achievement), a multilevel modelling approach was used, nested at the participant level, due to the existence of multiple observations per individual violating the assumption of independence of observation. Data for all 67 participants, for both intervention and control sessions were analysed, using an intervention session binary variable to control for intervention attributable effect on achievement. Subgroups (ID, ASC, ID/ASC) were examined separately.

The dependent variable was participant achievement and exposure variables were age, gender, affective state, experimental condition, subgroup and level of ID. Because the proportions of time spent in each of the three affective states sum to unity, only two of these proportions are independent. As such, one proportion must be excluded. We considered it prudent to exclude each in turn. A model was constructed for each possible pairing of the three affective states (engaged and frustrated; engaged and bored; frustrated and bored). These proportions, alongside our measure of achievement, were transformed to the natural log due to their non-Gaussian distributions and to ease interpretation. In two instances, participants spent no time in the frustrated state, so these proportions were set to 0.01 to allow conversion to the natural log.

Linear mixed and log-linear mixed models (Hox, Moerbeek, & Van de Schoot, 2017) were initially tested. Our inspection of fixed effect residuals did not indicate heteroskedasticity, yet, we observed significant clustering around the achievement ceiling which may have led to underestimation of regression coefficients. To control for this ceiling effect, one common to the measure of educational outcomes (Wang, Zhang, McArdle, & Salthouse, 2009), we adopted a multilevel mixed effects tobit model (Barros, Galea, Leiva, & Santos-Neto, 2018). Model fitting and variable selection utilised the Akaike Information Criterion (AIC) (Bozdogan, 1987), which performs better than the Bayesian Information Criterion (BIC) in the case of noncomplex linear models with disparate observations within nests (Vrieze, 2012).

Age and gender did little to improve the fit of the models, but were included to control for sampling population heterogeneity given the experimental study design. Level of ID, as a more granular measure, improved model fit compared to membership of subgroup. This was because there was a significant correlation between the two, that is, those designated severely disabled were found predominantly in the joint ID/ASC subgroup. To characterise individual propensity to achieve and individual-variation in the effects of affective states on achievement, each model allowed a participant-specific random intercept and a random slope for the natural log of engaged in models A&B and of bored in model C. It was not possible to fit a random coefficient to the affective state of frustration, probably because of the low values obtained for that affective state. An independent covariance structure was specified as all examined covariances did not differ significantly from zero.

Final model specification took the forms:

$$\begin{aligned} \text{A. } \ln\text{achievement}_{ij} = & \beta_0 + \beta_2 \text{age}_{ij} + \beta_3 \text{female}_{ij} + \beta_4 \ln\text{engaged}_{ij} + \beta_5 \ln\text{frustrated}_{ij} \\ & + \beta_6 \text{intervention}_{ij} + \beta_7 \text{mild}_{ij} + \beta_8 \text{moderate}_{ij} + \beta_9 \text{severe}_{ij} \\ & + u_{i0} + u_{i1} \ln\text{engaged}_{ij} + \varepsilon_{ij}. \end{aligned}$$

$$\begin{aligned} \text{B. } \ln\text{achievement}_{ij} = & \beta_0 + \beta_2 \text{age}_{ij} + \beta_3 \text{female}_{ij} + \beta_4 \ln\text{engaged}_{ij} + \beta_5 \ln\text{bored}_{ij} \\ & + \beta_6 \text{intervention}_{ij} + \beta_7 \text{mild}_{ij} + \beta_8 \text{moderate}_{ij} + \beta_9 \text{severe}_{ij} \\ & + u_{i0} + u_{i1} \ln\text{engaged}_{ij} + \varepsilon_{ij} \end{aligned}$$

$$\begin{aligned}
 \text{C. } \ln\text{achievement}_{ij} = & \beta_0 + \beta_2 \text{age}_{ij} + \beta_3 \text{female}_{ij} + \beta_4 \ln\text{frustrated}_{ij} + \beta_5 \ln\text{bored}_{ij} \\
 & + \beta_6 \text{intervention}_{ij} + \beta_7 \text{mild}_{ij} + \beta_8 \text{moderate}_{ij} + \beta_9 \text{severe}_{ij} \\
 & + u_{i0} + u_{i1} \ln\text{bored}_{ij} + \varepsilon_{ij}
 \end{aligned}$$

where the dependent variable  $\ln\text{achievement}$  represents the natural log of achievement of participant  $i$  in session  $j$ . The independent variables of age and the natural log of proportion of time spent in affective states were coded as continuous variables. Levels of ID and being female were specified as binary variables.

Wald  $z$  tests were performed under the central limit theorem to test the individual significance of fixed effect coefficients (Bolker *et al.*, 2009). However, resultant  $p$  values should be taken conservatively given the experimental and subjective study design (Johansson, 2011).

In order to address the second hypothesis (MaTHiSiS had a positive effect on engagement and learning achievement), a mean score for engagement and for achievement was calculated for each participant. As these data met the requirements for parametric analysis, a related  $t$  test was used to compare scores between the two conditions.

## Results

**Hypothesis 1:** The sensor data can automatically identify different affective states associated with learning achievement.

Results from the three different models for the whole group ( $N = 67$ ) are shown in Table 2.

While models B and C performed equally well, model A including both proportion engaged and frustrated fit the data best as judged by the Akaike Information Criterion. In model A, both engagement and frustration are positively associated with achievement: for every 1% increase in the proportion of time spent engaged there is an increase of 0.475% ( $p < 0.0001$ ) in achievement and similarly for frustration an increase of 0.128% ( $p < 0.0001$ ). There exists a significant random effect between participants on achievement of the proportion of time spent engaged. In the second model (B), a 1% increase in proportion of time bored is associated with a reduction in achievement of 0.188% ( $p < 0.0001$ ) and in the third model by 0.201% ( $p < 0.0001$ ). In all three models, being described as having a severe disability is negatively associated with achievement, reducing achievement by 24.86% ( $p < 0.0001$ ) in the first model, 20.09% in the second and 22.38% in the third, compared to not having any ID. Neither age, nor gender, contribute significantly to achievement in any of the models. In all three models, the intervention is associated with increases in participant achievement ranging from 1.5% (A) to 3.7% (B), although these were not statistically significant.

Subgroup analysis did not demonstrate any divergence from the findings of our primary analysis.

**Hypothesis 2:** MaTHiSiS has a positive effect on engagement and learning achievement.

Means and standard deviations for the total sample and three sub groups are shown in Table 3. While the previously discussed models A, B and C (see Table 2) showed a positive but not significant effect of intervention on achievement, when means of intervention sessions are compared for the group as a whole, there was a significantly ( $t = 3.769$ ,  $df = 66$ ,  $p < 0.0004$ , Cohen's  $d = 0.460$ ) higher proportion of the session spent being engaged and a significantly lower proportion of the session spent being bored ( $t = 3.852$ ,  $df = 66$ ,  $p < 0.0003$ , Cohen's  $d = 0.471$ ), in the

Table 2: Multilevel mixed-effects tobit regressions of natural log achievement ( $n = 649$ )

Model	Variable	$b$ (95% CI)	$p$	Random effect
A	Constant	0.909 (0.682 to 1.136)	0.000	0.117
	Age	-0.002 (-0.02 to 0.015)	0.800	-
	Female	0.06 (-0.03 to 0.15)	0.189	-
	ln(engaged)	0.475 (0.36 to 0.591)	0.000	0.238
	ln(frustrated)	0.128 (0.088 to 0.167)	0.000	-
	Intervention	0.015 (-0.026 to 0.056)	0.470	-
	Mild disability	0.004 (-0.131 to 0.139)	0.955	-
	Moderate disability	-0.047 (-0.171 to 0.078)	0.461	-
B	Severe disability	-0.286 (-0.404 to -0.168)	0.000	-
	Constant	0.102 (-0.152 to 0.357)	0.431	0.086
	Age	-0.0002 (-0.018 to 0.017)	0.975	-
	Female	0.031 (-0.059 to 0.122)	0.500	-
	ln(engaged)	0.072 (-0.08 to 0.224)	0.355	0.245
	ln(bored)	-0.188 (-0.251 to -0.125)	0.000	-
	Intervention	0.036 (-0.005 to 0.076)	0.082	-
	Mild disability	0.003 (-0.122 to 0.128)	0.960	-
C	Moderate disability	-0.051 (-0.166 to 0.064)	0.386	-
	Severe disability	-0.224 (-0.353 to -0.096)	0.001	-
	Constant	0.183 (-0.052 to 0.419)	0.127	0.161
	Age	-0.003 (-0.022 to 0.015)	0.707	-
	Female	0.042 (-0.058 to 0.141)	0.412	-
	ln(frustrated)	0.041 (0.002 to 0.08)	0.041	-
	ln(bored)	-0.201 (-0.248 to -0.155)	0.000	0.014
	Intervention	0.021 (-0.021 to 0.064)	0.322	-
	Mild disability	0.009 (-0.129 to 0.148)	0.894	-
	Moderate disability	-0.081 (-0.208 to 0.046)	0.210	-
	Severe disability	-0.253 (-0.388 to -0.119)	0.000	-

$n$  = the total number of participant sessions. ln(affective state) represents the natural log of the proportion of time spent in the designated affective state. Coefficients not in natural log form can be transformed to percentage change on achievement using  $(e^{\text{coefficient}} - 1) \times 100$ .

intervention condition than in the control condition. However, there was no significant difference between the proportion of the session spent being frustrated or the achievement scores for the two conditions.

This pattern of results was repeated for the participants with ID and for those with ID and ASC. For those with ID there was a significantly ( $t = 2.924$ ,  $df = 22$ ,  $p < 0.008$ , Cohen's  $d = 0.610$ ) higher proportion of the session spent engaged and a significantly lower proportion of the session spent bored ( $t = 3.945$ ,  $df = 2$ ,  $p < 0.001$ , Cohen's  $d = 0.822$ ) in the intervention condition than in the control condition. For the participants with ID and ASC there was a significantly ( $t = 2.843$ ,  $df = 21$ ,  $p < 0.01$ , Cohen's  $d = 0.485$ ) higher proportion of time spent being engaged and a significantly lower proportion of time spent bored ( $t = 3.945$ ,  $df = 22$ ,  $p < 0.001$ , Cohen's  $d = 0.606$ ) in the intervention condition than in the control condition. For both groups, there was no significant difference between the proportion of the session spent being frustrated or the achievement scores for the two conditions (although achievement is higher for these two groups in the intervention condition).

For the participants with ASC only, although the mean proportion of the session spent being engaged was higher in the intervention condition, this difference did not reach significance.

Table 3: Means and standard deviations for the complete group and subgroups

Use case (N)	A—Intervention				B—Control			
	Engaged	Bored	Frustrated	Achievement	Engaged	Bored	Frustrated	Achievement
Intellectual disabilities (23)	0.591 (0.105)	0.164 (0.065)	0.248 (0.094)	1.511 (0.246)	0.527 (0.052)	0.223 (0.052)	0.255 (0.083)	1.502 (0.291)
ASC only (22)	0.589 (0.084)	0.166 (0.055)	0.243 (0.080)	1.517 (0.273)	0.559 (0.102)	0.180 (0.072)	0.262 (0.137)	1.548 (0.299)
Intellectual disabilities and ASC (22)	0.549 (0.083)	0.164 (0.043)	0.287 (0.069)	1.306 (0.347)	0.509 (0.049)	0.187 (0.026)	0.301 (0.051)	1.249 (0.366)
Total (67)	0.577 (0.092)	0.165 (0.054)	0.259 (0.083)	1.446 (0.303)	0.532 (0.074)	0.197 (0.056)	0.272 (0.098)	1.434 (0.342)

There was no difference between the proportion of the session spent being bored or frustrated or in the achievement scores for the two conditions.

In order to determine whether the effect on achievement was different depending on length of exposure, participants with a duration less than 60 minutes were excluded from the analysis. However, there was still no significant difference between achievement scores from intervention and control sessions for the group as a whole or for each subgroup.

## Discussion

The first hypothesis is accepted as results from the multilevel model indicate that the sensor data can identify three different affective states all with a strong relationship with achievement irrespective of experimental condition. The state labelled “lack of boredom” is the state most strongly linked to achievement, whilst those labelled “frustration” and “engagement” are positively related to achievement. These conclusions are supported by similar findings from the modelling of the data from each of the three subgroups (ID, ID/ASC, ASC) separately, indicating the robustness of the final model.

When interpreting these results, it has to be emphasised that these variables represent how the system was interpreting the affective state of the learner. The initial ground truth exercise was carried out on a limited database compared with other studies and there was no opportunity to collect teacher rated affective states with which to compare automatically detected states as performed by Grawemeyer *et al.* (2017). During the intervention sessions, the detection of affective states was used to manipulate the presentation of learning material. Thus, a relationship between automatically detected affective state and achievement was to be expected. However, the multilevel model indicated that the relationships between the different affective states and achievement held regardless of type of session, even when the system was not deliberately linking its calculation of affective state to changing the presentation of material. This suggests that the MaTHiSiS algorithms are identifying affective states that are independently linked to achievement although how they correspond to human detected states of engagement, frustration and boredom is not known.

The significance of the negative relationship between boredom and learning achievement is in line with previous tests of the salience of this affective state (Craig, Graesser, Sullins, & Gholson, 2004). The positive contribution of frustration is not inconsistent with the affective model proposed by D’Mello and Graesser (2012b). They proposed that the experience of cognitive disequilibrium when the learner is confronted with a contradiction, anomaly, system breakdown or error, could lead to being frustrated or stuck when the learner is uncertain about what to do next. This explained the transition into disengagement (boredom) which would reduce learning. However, it is only if this frustration is persistent that it would turn into boredom, when the learner disengages from the learning process (p. 147). It is, therefore, possible that the frustration detected by the MaTHiSiS system was of the transient kind either because the software adjusted to move the learner to a different state (by reducing the level of difficulty or by choosing alternative learning materials) or because the learner marshalled their own resources to meet the challenge that led to their frustrated state.

It is also possible that the MaTHiSiS system may have experienced difficulties in distinguishing between the states of frustration and engagement, due to both states sharing some of the features utilised by the affective state recognition (eg, eye gaze). The number of modalities used in this study could be supplemented with additional information from physiological measures and conversational cues (D’Mello, Dieterle, & Duckworth, 2017).

A potential limitation was the method of summarising the data for analysis, possibly masking the identification of transient affective states. D'Mello and Graesser (2012b) make the point that overall measures of affective state during an entire learning session are unsatisfactory because learners oscillate between positive and negative states throughout a session. More interesting is "a fine-grained analysis of the rapid dynamics of both positive and negative affective states that naturally occur during effortful learning activities" (p. 146) an approach adopted by the *italk-2learn* platform (Grawemeyer *et al.*, 2017). These data are available from the MongoDB database and a future examination of moment by moment states is possible.

The salience of the three affective states also described the situation for the three subgroups (ID, ID/ASC and ASC only). A characteristic of the learner that was more important was level of ID, with those with severe ID demonstrating significantly lower levels of achievement whatever model was used. This suggests that it would be beneficial to re-examine the suitability and limited range of the learning materials provided.

The second hypothesis is partially met as when the two sets of sessions are compared on levels of the different affective states, the responsive nature of the system did increase the proportion of time learners were engaged at the expense of boredom. Disappointingly, it did not produce significantly greater achievement in terms of correct responses to presented learning material (although achievement is higher in the ID and ID/ASC groups when using the intervention). This conclusion is echoed in the findings of the multilevel model. Although this relationship held even when participants with low exposure times were omitted, a further limitation of the study was that even greater levels of exposure may be required for the affect sensitive version of the MaTHiSiS system to be obvious. While attempting to increase the responsiveness of the system to the learners' changing affective states may have maintained the higher levels of engagement, the rapid responsiveness, for example in response to frustration, may have moved them out of this state too early for learners to employ the thought, reflection and problem solving necessary to increase achievement (D'Mello & Graesser, 2012b).

An improvement in affective states but not performance was also reported by Aslan *et al.* (2019) who found that, although their intervention reduced boredom, the improvement pre to posttest in the intervention group was not significantly higher than that in the control group. When evaluating the addition of "human provided emotional scaffolding" to an automated reading tutor, Aist, Kort, Reilly, Mostow, and Picard (2002) observed improved persistence with the task, but no improvement in students' memory of facts. The crucial role of persistence was also highlighted by Nakamura and Csikszentmihalyi (2009), who reported that in formal learning situations flow experiences predict greater persistence and achievement in the associated activity over the long term. This reinforces the conclusion of Thompson and McGill (2017) that the benefits of an affective tutoring system may be more apparent in the longer term, rather than after a short evaluation. When considering the lack of a significant difference in achievement between the two conditions in the present study, it is worth remembering that the control condition itself is an intelligent tutoring system, but one based on achievement alone. It would be surprising if learners did not show a reasonable level of achievement during control sessions and that the benefits of including affective sensing would only become significant over a longer period of exposure. Increased engagement and decreased boredom in the affect driven version of the MaTHiSiS system suggest that learners with ID may be more likely to invest in the time and number of repetitions required for this achievement to become evident.

The study also suffered from a limited selection of learning material which may have been too restricted to demonstrate much variation in outcomes. A ceiling effect was also proposed by Thompson and McGill (2017) to explain the lack of learning in their study of an affective tutoring

system to teach genetics. They used a similar control to that of the present study and found significantly greater levels of perceived learning in their intervention group, but no differences in levels of content knowledge as measured using a summary quiz. They proposed using more challenging materials and this is consistent with findings from other studies (D'Mello & Graesser, 2012a).

While the relationship between affective states and achievement held whatever subgroup the learners were allocated to, the comparison between the affect sensitive version of the MaTHiSiS system and the control revealed a slightly different result for the participants having predominantly an ASC. Unlike the ID and ID/ASC groups, although the mean proportion of time spent being engaged was higher in the intervention condition, this difference did not reach significance. It has been hypothesised that as a result of difficulties regulating their arousal levels, people with autism show atypical attention, specifically attention reorienting difficulties not just with social stimuli (Orekhova & Stroganova, 2014). Therefore, reliance on the five modalities used in this study may be less helpful for this group of learners. Other approaches to detecting affective state in students with autism have prioritised the use of accelerometer data (Sumi *et al.*, 2018). Caution should be exercised in the conclusions to be drawn here, as sites in which testing occurred may have used different approaches to determine presence of an ASC. However, it does highlight the need to investigate whether affect detection and its precise role in determining the presentation of learning material, may need to take a different form for learners with autism.

## Conclusions

This is the first study to evaluate an adaptive learning system for learners with ID based on multi-modal affect recognition. Three separate states were automatically identified, with lower levels of the state labelled “boredom” having the strongest link to learning achievement. Both those labelled “frustration” and “engagement” were positively related to achievement. Our results are in line with other studies showing that engagement increases when activities are tailored to the personal needs and emotional states of learners (Athanasiadis *et al.*, 2017), but no significant difference in learning achievement was found (at least for the period of our study) when adaption was based on both the affective state and achievement of the learner, compared with achievement alone.

Although future work is necessary to refine the machine learning methods and develop a greater range of learning material, an adaptive learning system based on affect recognition that can support teachers of those with ID is one step nearer. Such a system that responds in real time to learners' affective states allows teachers to decide on a more appropriate distribution of their close monitoring to ensure they can target their efforts so that all students are supported to reach their full potential.

## Acknowledgements

This work was supported by the Horizon 2020 fund (MaTHiSiS—Managing Affective-learning THrough Intelligent atoms and Smart InteractionS—S687772). The authors would like to thank all the schools who collaborated in data collection and Joe Sarsfield for comments on the manuscript.

## Statements on open data, ethics and conflict of interest

The file containing anonymised data on which the analysis was performed is available from the corresponding author. This research was conducted under approval from the University of Nottingham, Faculty of Medicine and Health Sciences Research Ethics Committee.

Teachers gave informed consent and parents or guardians gave consent on behalf of their children before data collection.

No authors have any conflicts of interest to declare with regard to the publishing of this work.

## References

- Aist, G., Kort, B., Reilly, R., Mostow, J., & Picard, R. (2002, October). *Experimentally augmenting an intelligent tutoring system with human-supplied capabilities: Adding human-provided emotional scaffolding to an automated reading tutor that listens*. In Presentation at the 4th IEEE International Conference on Multimodal Interfaces (pp. 483–490). Pittsburgh, PA.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., D'Mello, S. K., & Esme, A. A. (2019). *Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms*. Presentation at Computer Human Interaction (CHI) 2019 (paper 304), Glasgow, Scotland.
- Athanasiadis, C., Hortal, E., Koutsoukos, D., Lens, C. Z., & Asteriadis, S. (2017). Personalized, affect and performance-driven computer-based learning. *Proceedings of the 9th International Conference on Computer Supported Education—Volume 1: CSEDU*. Porto, Portugal, April 21–23, 2017.
- Baker, R. S. J., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223–241.
- Barros, M., Galea, M., Leiva, V., & Santos-Neto, M. (2018). Generalized Tobit models: Diagnostics and application in econometrics. *Journal of Applied Statistics*, 45(1), 145–167.
- Basawapatna, A. R., Repenning, A., Han Koh, K., & Nickerson, H. (2013, August). The zones of proximal flow: Guiding students through a space of computational thinking skills and challenges. In *Proceedings of the ninth annual international ACM conference on International computing education research* (pp. 67–74). San Diego, CA.
- Bertini, E. S., & Kimani, S. (2003). Mobile devices: Opportunities for users with special needs: Mobile HCI. In L. Chittaro (Ed.), *Lecture notes of computer science* (Vol. 2795, pp. 486–491). Udine, Italy: Springer Verlag.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Henry, M., ... White, J. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135.
- Boulton, H., Brown, D., Standen, P., Belmonte, M., Kwiatkowska, G., Hughes-Roberts, T., & Taheri, M. (2018, March 5–7). Multi-modalities in classroom learning environments. *INTED2018 Proceedings: 12th Annual International Technology, Education and Development Conference*. Valencia, Spain: IATED Academy.
- Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345–370.
- Chickerur, S., & Joshi, K. (2015). 3D face model dataset: Automatic detection of facial expressions and emotions for educational environments. *British Journal of Educational Technology*, 46(5), 1028–1037.
- Colley, A. (2013). *Personalised learning for young people with profound and multiple learning difficulties*. London, UK: Jessica Kingsley.
- Craig, S., Graesser, A., Sullins, J., & Gholson, J. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media*, 29, 241–250.
- Cukurova, M., Kent, C., & Luckin, R. (2019). Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology*, 50(6), 3032–3046.
- D'Mello, S., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modelling and User Adapted Interaction*, 20(2), 147–187.
- D'Mello, S., & Graesser, A. (2012a, December). AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiIS)—Special issue on highlights of the decade in interactive intelligent systems*, 2(4), Article No. 23.



- D'Mello, S., & Graesser, A. (2012b). Dynamics of affective states during complex learning. *Learning and Instruction*, 22, 145–157.
- D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) measurement of engagement during learning. *Educational Psychologist*, 52(2), 104–123.
- du Boulay, B., Poulouvasillis, A., Holmes, W., & Mavrikis, M. (2018). Artificial intelligence and big data technologies to close the achievement gap. In R. Luckin (Ed.), *Enhancing learning and teaching with technology* (pp. 256–285). London, UK: UCL Institute of Education Press.
- Ghaleb, E., Hortal, E., Koutsoukos, D., Amestoy, M., Spyrou, V., Sgouropoulos, D., ... Taheri, M. (2017a). Affect understanding in MaTHiSiS. Retrieved from <http://mathisis-project.eu/en/deliverables>
- Ghaleb, E., Popa, M., Hortal, E., & Asteriadis, S. (2017b). Multimodal fusion based on information gain for emotion recognition in the wild. In *2017 Intelligent Systems Conference (IntelliSys)* (pp. 814–823). London, UK: IEEE.
- Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., & Rummel, N. (2017). Affective learning: Improving engagement and enhancing learning with affect-aware feedback. *User Modeling and User-Adapted Interaction*, 27(1), 119–158.
- Hamari, J., Sherno, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, 170–179.
- Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.
- Johansson, T. (2011). Hail the impossible: P-values, evidence, and likelihood. *Scandinavian Journal of Psychology*, 52(2), 113–125.
- Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of the interplay between emotions and learning. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT '01)* (pp. 43–46). Madison, WI: IEEE Computer Society.
- Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H. E., Coulter, D. L., Craig, E. M. (P.), Reeve, A., ... The AAMR AD HOC Committee on Terminology and Classification. (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). Washington, DC: American Association on Mental Retardation.
- Mazzucato, A., & Traversi, M. (2017). D8.5 Report on Profound and Multiple Learning Disabilities Case pilots. Retrieved from <http://mathisis-project.eu/en/content/d85-report-profound-and-multiple-learning-disabilities-case-pilots-m24>
- Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow optimal experience and its role in development. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). New York, NY: Oxford University Press.
- Nakamura, J., & Csikszentmihalyi, M. (2009). Flow theory and research. *Handbook of positive psychology*. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 195–206). New York, NY: Oxford University Press.
- Nesta. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. Retrieved from <https://www.nesta.org.uk/report/education-rebooted/>
- Nganji, J., & Brayshaw, M. (2017). Disability-aware adaptive and personalised learning for students with multiple disabilities. *International Journal of Information and Learning Technology*, 34(4), 307–321. <https://doi.org/10.1108/IJILT-08-2016-0027>
- Ocuppaugh, J., Baker, R. S. J., & Rodrigo, M. M. T. (2012). *Baker-Rodrigo Observation Method Protocol (BROMP) 1.0. Training Manual Version 1.0* (Technical report). New York, NY: EdLab.
- Orekhova, E., & Stroganova, T. (2014). Arousal and attention re-orienting in autism spectrum disorders: Evidence from auditory event-related potentials. *Frontiers in Human Neuroscience*, 8, 34.
- Porayska-Pomsta, K., Alcorn, A. M., Avramides, K., Beale, S., Bernardini, S., Foster, M. E., ... Smith, T. J. (2018, December). Blending human and artificial intelligence to support autistic children's social communication skills. *ACM Transactions on Computer-Human Interaction*, 25(6), Article 35, 35 pages. <https://doi.org/10.1145/3271484>.
- Rose, R. M., & Blomeyer, R. L. (2007). Access and equity in online classes and virtual schools. *North American Council for Online Learning*. Retrieved from <https://www.inacol.org/resource/access-and-equity-in-online-classes-and-virtual-schools>

- Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., ... Yeager, M. H. (2010). *Intellectual Disability: Definition, Classification, and Systems of Supports*. (11th ed.). Washington, DC: American Association on Intellectual and Developmental Disabilities. <http://www.aaid.org/ISBN-978-1-9353-0404-3>
- Scheiter, K., Schubert, C., Schuler, A., Schmidt, H., Zimmermann, G., Wassermann, B., ... Edera, T. (2019). Adaptive multimedia: Using gaze-contingent instructional guidance to provide personalized processing support. *Computers & Education*, 139, 31–47.
- Sumi, A. I., Zohora, M. F., Mahjabeen, M., Faria, T. J., Mahmud, M., & Kaiser, M. S. (2018). fASSERT: A fuzzy assistive system for children with autism using internet of things. In *Brain informatics, lecture notes in computer science* (pp. 403–412). Arlington, TX: Springer International Publishing.
- Taheri, M., Brown, D., Standen, P., Sherkat, N., Cosma, G., & Langensiepen, C. (2018, September 4–6). State diagram for affective learning in an educational platform. *Proc. 12th ICDVRAT with ITAG*. Nottingham, UK.
- Taub, D. A., McCord, J. A., & Ryndak, D. L. (2017). Opportunities to learn for students with extensive support needs: A context of research-supported practices for all in general education classes. *The Journal of Special Education*, 51(3), 127–137.
- Thompson, N., & McGill, T. J. (2017). Genetics with Jean: The design, development and evaluation of an affective tutoring system. *Education Technology Research and Development*, 65, 279–299.
- Tsatsou, D., Pomazanskyi, A., Hortal, E., Spyrou, E., Leligou, H., Asteriadis, S., ... Daras, P. (2018). Adaptive learning based on affect sensing. *Artificial Intelligence in Education, AIED 2018. Lecture Notes in Computer Science*, 10948.
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 17(2), 228–243.
- Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2009). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research*, 43(3), 476–496.
- Yadegaridehkordi, E., Noor, N. F. B. M., Ayub, N. M. B. A., Affal, H. B., & Hussin, N. B. (2019, December). Affective computing in education: A systematic review and future research. *Computers & Education*, 142, 103649. <https://doi.org/10.1016/j.compedu.2019.103649>