International Journal of Artificial Intelligence in Education 18 (2013) xxx – xxx
IOS Press

Towards a Framework for Modelling Engagement Dynamics in Multiple Learning Domains

Tanja Käser[1], Gian-Marco Baschera[1], Alberto Giovanni Busetto[1,2], Severin Klingler[1], Barbara Solenthaler[1], Joachim M. Buhmann[1,2], and Markus Gross[1]
[1]*Department of Computer Science, ETH Zurich, Zurich, Switzerland*
[2]*Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland*
E-Mail: kaesert@inf.ethz.ch

Abstract. In this paper, we explore the possibility of a general framework for modelling engagement dynamics in software tutoring, focusing on the cases of developmental dyslexia and developmental dyscalculia. This project aims at capturing the similar engagement state patterns for the two learning disabilities. We start by presenting a model of engagement dynamics in spelling learning, which relates input behaviour to learning and explains the dynamics of engagement states. Predictive power of extracted features is increased by incorporating domain knowledge in the pre-processing. The introduced model enables the prediction of focused and receptive states, and of forgetting. In the second part, we extend the model to a more general framework, which takes into account the similarities and dissimilarities of the two studied cases. Finally, we define desirable properties of a general engagement dynamics model, while analysing the reusability of the introduced spelling model.

Keywords. engagement modelling, dyslexia, dyscalculia, dynamic Bayesian network, human learning, spelling, mathematics learning

INTRODUCTION

Affective modelling is receiving increasing attention due to its recognized relevance in learning. It is considered a particularly challenging task for two main reasons. First, ground truth is unattainable, and thus it invariably requires indirect measures and approximations. Second, experimental data are limited in quantity and quality due to high costs and significant noise levels. In our previous work (Baschera et al., 2011), we have developed an engagement dynamics model in spelling learning that can adapt the training to individual students based on data-driven identification of engagement states from student input. Building upon this model, we explore whether we can transfer the existing framework to a more general engagement dynamics model for multiple learning domains.

Related work
In general, affective models can be inferred from several sources, such as sensor data (Cooper et al., 2010; Heraz and Frasson, 2009) and user input data (Baker et al., 2004; Johns and Woolf, 2006; Arroyo and Woolf, 2005). These sources differ in what they measure as well as in how that measurement occurs. On the one hand, sensor measurements tend to be direct and comprehensive. They have the potential to measure larger numbers of affective features. On the other hand, input measurements are not limited to laboratory experimentation. The measurement of student interaction with a software tutoring system offers a unique opportunity: Large and well-organized sample sets can be obtained from a variety of experimental conditions. These recorded inputs exhibit the potential to characterize the affective state of the student in a learning scenario. It has been shown that highly informative features, such as problem timing and hint requests, can be extracted from log files (Arroyo and Woolf, 2005). The identification of informative features and the incorporation of domain knowledge, either as implicit or as explicit assumptions, can substantially increment the predictive power of the inferred models (Busetto et al., 2009). Recent advances in feature selection enable the optimization of experimental design to identify complex systems (Busetto et al., 2009). In affective modelling, median splitting (Arroyo and Woolf, 2005), thresholding (Johns and Woolf, 2006), and input averaging (Baker et al., 2004) are established techniques for pre-processing.

Contribution
This study explores the definition of a general framework for modelling engagement dynamics in human learning. In particular, we focus on developmental dyslexia and dyscalculia. We argue that the assumption of similar engagement patterns in the two cases is justified and, thus, that a similar engagement model would be beneficial. Starting from a model for engagement dynamics in spelling learning (Baschera et al., 2011), we extend the introduced framework to the more general case of engagement modelling. We provide a detailed assessment of similarities and dissimilarities of the two cases of developmental dyslexia and dyscalculia in terms of learning domain, student model, and available data. Furthermore, we analyse the reusability of the engagement model for spelling learning and define desirable properties of a general model of engagement dynamics for software tutoring.

COMPARISON OF LEARNING DISABILITIES

Developmental dyslexia and developmental dyscalculia are specific learning disabilities inferring a lack of success in language processing and mathematics, respectively. In this section, we discuss both learning disabilities and existing intervention programs. We highlight the similarities between the conditions, which indicate the possibility of similar engagement patterns.

Dyslexia

Developmental dyslexia is a specific learning disability which affects the acquisition of reading and writing skills (World Health Organization, 1993). Children with developmental dyslexia tend to exhibit inconsistent orthography speed and accuracy problems, as well as difficulty in segmenting and manipulating phonemes in words. In addition to poor writing and reading skills, poor speech production and spelling are other symptoms of developmental dyslexia (Goswami, 2003). Currently, developmental dyslexia is thought to originate from a neurological disorder with genetic origin (Galaburda et al., 1985, 2006; Schulte-Korne et al., 2004; Demonet et al., 2004; Ziegler et al., 2005). The prevalence of this disability is estimated to range from 5% to 17.5% in English speaking countries (Shaywitz, 1998), and to about 10% in German speaking countries (Russeler et al., 2006).

Intervention

There exist a lot of intervention programs to remediate developmental dyslexia that have been scientifically evaluated in children (and adults). These programs predominantly aim at training auditory and visual functions using approaches such as low-level auditory perceptual learning (Tallal, 2004; Robichon et al., 2002; Santos et al., 2007; Besson et al., 2007; Gaab et al., 2007; Uther et al., 2006), practice of speech-like auditory stimuli (O'Shaughnessy and Swanson, 2000; Hatcher et al., 2006), practice of specific manipulations of speech-like signals (Tallal, 2004), improvement of high- and low-level visual functions (Bacon et al., 2007; Lorusso et al., 2006) and combined training of auditory and visual functions (Kujala et al., 2001). Other intervention techniques combine the training of reading and writing skills (Vadasy et al., 2000; Edwards, 2003; Shaywitz et al., 2004). Lately, a few multi-modal training programs have been proposed as well (Kujala et al., 2001; Gross and Vögeli, 2007; Kast et al., 2007).

Dyscalculia

Developmental dyscalculia is a specific learning disability affecting the acquisition of arithmetic skills (von Aster and Shalev, 2007). Genetic, neurobiological, and epidemiological evidence indicates that developmental dyscalculia is a brain-based disorder, although poor teaching and environmental deprivation have also been discussed in its aetiology (Shalev, 2004). Developmental dyscalculia is thought to have its neuropsychological basis due to limited 'number sense', which implies a deficit in very basic numerical skills such as number comparison (Landerl et al., 2004; Rubinsten and Henik, 2005; Butterworth, 2005a,b). Besides exhibiting fundamental deficits in number processing (Cohen Kadosh et al., 2007; Mussolin et al., 2010; Kucian et al., 2006; Price et al., 2007), children with developmental dyscalculia also tend to suffer from difficulties in acquiring simple arithmetic procedures and exhibit a deficit in fact retrieval (Ostad, 1997, 1999). The prevalence of developmental dyscalculia is estimated to about 3-6% (Shalev and von Aster, 2008; Badian, 1983; Lewis et al., 1994) in English and German speaking countries.

Intervention

Several intervention programs for developmental dyscalculia have been proposed and the scientific evaluation in children proved overall successful. Existing interventions can be categorized according to the target age and the approaches used. Early intervention programs for young children 'at risk' of developing arithmetic difficulties focus mostly on training basic numerical skills (Griffin et al., 1994; Wright, 2003; Van De Rijt and Van Luit, 1998). Other interventions are individualized remedial programs for primary school children with difficulties in mathematics (Dowker, 2001; Kaufmann et al., 2003). There are intervention programs which specifically aim at training number representations (Kucian et al., 2011; Wilson et al., 2006a,b, 2009; Siegler and Ramani, 2009), while other learning programs align their content to the curriculum taught in school (Lenhard et al., 2011).

Comorbidity and similarities in engagement

Developmental dyslexia and developmental dyscalculia, both brain-based disorders, often exhibit comorbidity, which is the co-occurrence of two or more disorders in the same individual. Studies show that individuals with developmental dyscalculia do often show language difficulties as well, and vice versa, that dyslexic individuals often suffer from difficulties in arithmetic (von Aster and Shalev, 2007; Ostad, 1998; Lewis et al., 1994; Badian, 1999; Barbaresi et al., 2005; Dirks et al., 2008; Ackerman and Dykman, 1995). More importantly, children with these learning disabilities often also show comorbidities with ADHD (Shaywitz et al., 1994; Germanò et al., 2010; Fletcher, 2005; Barbaresi et al., 2005). In addition, these learning disabilities often lead to anxiety and aversion against the subject (Rubinsten and Tannock, 2010) and to underperformance in school and later in profession (Bynner, 1997). These facts suggest that children with learning disabilities will exhibit low intrinsic motivation and attentional problems and thus, monitoring of engagement dynamics becomes even more important. Since similar implications are relevant for the two learning disabilities, we assume the appearance of similar engagement states for developmental dyslexia and developmental dyscalculia.

LEARNING ENVIRONMENTS

The modelling of engagement dynamics is highly dependent on the properties of the learning environment it will be applied to. In this study, we investigate two different learning environments: the first one is a training program for spelling learning, *Dybuster*, and the second one, *Calcularis*, is intended for mathematics learning.

Dybuster

Dybuster (Gross and Vögeli, 2007; Kast et al., 2007) is a multi-modal training program for spelling learning. The central idea of the training software is to recode a sequential textual input string into a multi-modal representation using a set of codes. These codes reroute textual information through multiple undistorted visual and auditory cues. This training strategy builds up the memory strength of graphemes and phonemes. Visual cues include colours, forms and topology. Based on the information theoretical model of Dybuster, eight different colours are used in the software. The mapping of letters to colours is the result of a multi-objective optimization. For example, letters easily confused by dyslexics, e.g., ‘m’ and ‘n’, map to visually distinct colours. The idea is to associate colours with letters to eliminate mistakes due to letter confusion. The shapes are: spheres for small letters, cylinders for capital letters, and pyramids for the umlauts. The graph structure finally shows the decomposition of a word into syllables and graphemes. An additional auditory code computes a word-specific melody that is played to the user when entering a word. The different codes not only transfer information, but also stimulate different senses. This multi-sensory stimulation enhances perception and facilitates the retrieval of memory (Lehmann and Murray, 2005; Shams and Seitz, 2008).

Dybuster consists of three different games. In the COLOUR game (Fig. 1, top left), children learn the associations between colours and letters. Children need to remember the colours of the different letters: The colour fades out over time and children need to pick the right one. In the GRAPH game, children graphically segment a word into its syllables and letters (Fig. 1, top right). These first two games are played at the beginning of the training to learn the codes that are integrated in Dybuster. In the third game WORD LEARNING, representing the actual learning game, the program presents the alternative representations (graph, colours, shapes) of a word (Fig. 1, center). A voice dictates a word and the children hear a melody computed from the involved letters and the lengths of the syllables. Children then need to type the word on the keyboard. To avoid displaying completely misspelled words, the training program provides immediate visual and auditory feedback to errors. The sequence of words presented to the child is adapted to the skill level and the error profile of the children.

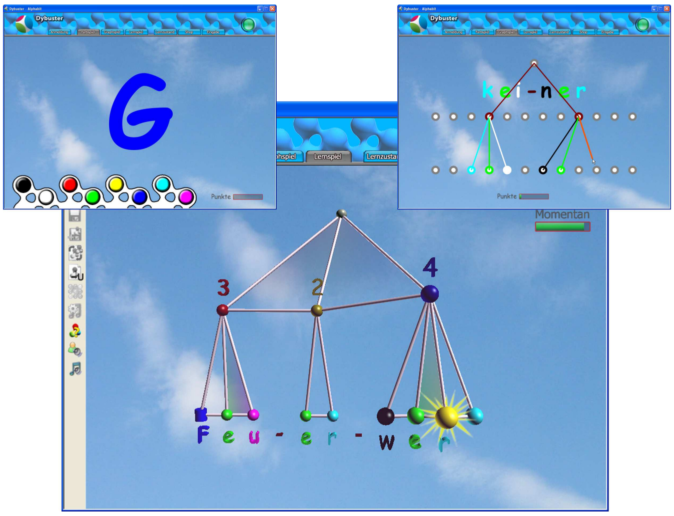


Fig.1. The three learning games of Dybuster: COLOUR game to train the associations between colours and letters (top left), GRAPH game for the training of the syllable structure (top right) and WORD LEARNING game with visual presentation of the different cues (bottom center).

Training motivation is an important component for the efficacy of a learning program. The learning environment of Dybuster features 3D graphics and interaction components and thus allows immersion in a playful 3D world. The different information channels interact with the user and give auditory and visual feedback if errors occur. Additionally, children collect points during the training that can be converted to virtual money. With this money, visual and auditory effects can be bought (Fig. 2).

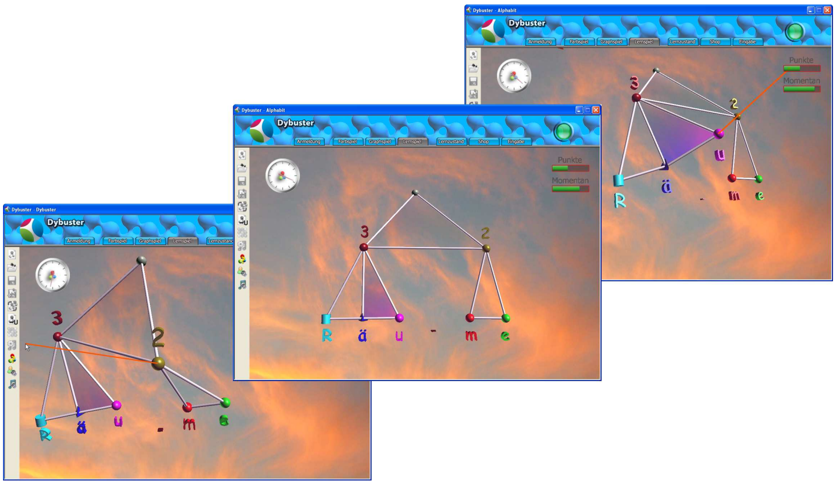
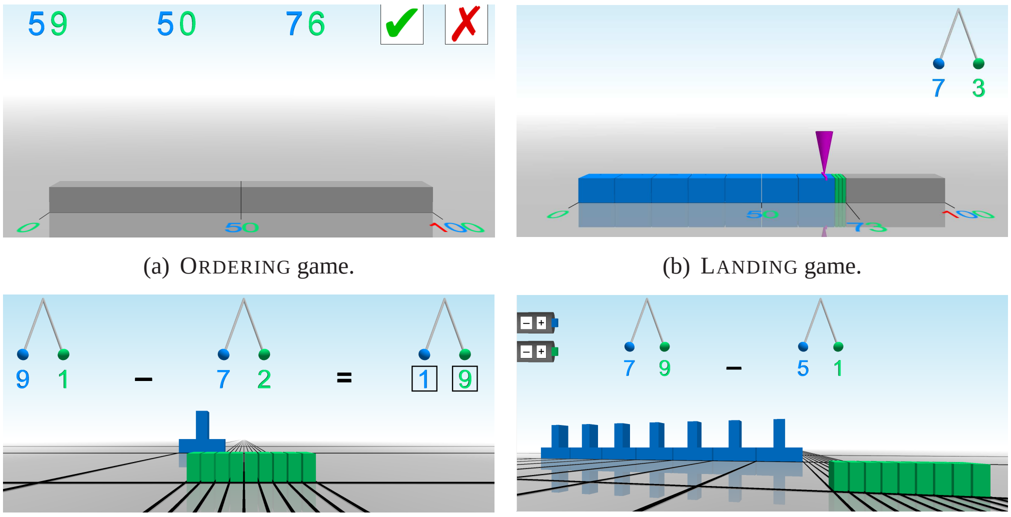


Fig.2. Example of visual interaction effects that can be bought in the shop.

Calcularis

Calcularis (Käser et al., 2011, 2012) is a training program for children with developmental dyscalculia or difficulties in learning mathematics. It consists of multiple games in a hierarchical structure. The games are structured according to number ranges and can be divided into two areas. The first area focuses on number representation and number understanding in general. Games in this area train the transcoding between different number representations and introduce the three principles of number understanding: cardinality, ordinality and relativity. The transcoding games are ordered according to the 'four-step developmental model' (von Aster and Shalev, 2007): Starting from a (probably) inherited core-system representation of cardinal magnitude (step 1), the linguistic symbolization (spoken number) develops during pre-school time (step 2). The arabic symbolization (written number) is then taught in school (step 3) and finally the analogue magnitude representation (number as a position on a number line) develops (step 4). A typical game of this area is the ORDERING game (Fig. 3(a)), where children need to decide if a sequence of numbers is sorted in ascending order. Another important game is the LANDING game (Fig. 3(b)): The position of a given number needs to be indicated on a number line by steering a falling cone using a joystick. The second area focuses on addition and subtraction. Games in this area train mental addition and subtraction at different difficulty levels. The difficulty of the game is determined by the magnitude of the numbers included in the task as well as by task complexity. Furthermore, the area also trains the understanding of the according operation. An example game in this area is the CALCULATOR game (Fig. 3(c)), where children perform mental addition or subtraction. The result of the task is entered using the keyboard. Another important game is the PLUS-MINUS game (Fig. 3(d)), where an addition (subtraction) task needs to be modelled using one, ten and hundred blocks. Blocks can be added and removed with mouse clicks.

The learning environment is fully adaptive, game selection and task difficulty are adapted to the child and systematic errors can be recognized.



(a) ORDERING game. (b) LANDING game.

(c) CALCULATOR game. (d) PLUS-MINUS game.

Fig.3. The games of Calcularis: In the CALCULATOR game, the solution of the given task needs to be typed. In the PLUS-MINUS game, the task displayed needs to be modelled with blocks of tens and ones. Blocks can be added and removed by clicking on the buttons with + and - signs. In the ORDERING game, children need to decide if the given numbers are sorted in ascending order. In the LANDING game, the position of a given number (73) needs to be indicated on the number line using a joystick.

Calcularis uses a multi-modal approach applying a special design for numerical stimuli. The different properties of numbers are encoded with visual cues such as colour, form and topology. The program uses different colours for the positions of the place-value system and the digits of a number are attached to the branches of a graph. Moreover, numbers are represented as a composition of blocks with different colours indicating hundred, tens or individual units. Lastly, the position of a number is displayed on a number line. All the different stimuli are consistently used in each game of the software to reinforce links between different number representations and improve number understanding.

Comparison

Computer-assisted learning has become popular also for learning disabilities. The effectiveness of such systems for children with dyslexia or developmental dyscalculia has been demonstrated in recent user studies (Gross and Vögeli, 2007; Kast et al., 2007; Baschera and Gross, 2010; Käser et al., 2012; Kucian et al., 2011; Wilson et al., 2006a). The computer can be a valuable tool for teaching children with learning disabilities when adhering to the following three principles:

1. The content and the goals of the training program should be theory-based. It should consider how learning works in the particular domain and take into account the specific problems of children

with learning disabilities. The content of Calcularis is aligned to the natural development of mathematical abilities. Furthermore, a bug library enables recognition of typical errors. Dybuster uses a multi-sensory representation with cues adapted to the specific difficulties of dyslexic children.

2. The learning environment should be motivating and encouraging. This is particularly important for children with learning disabilities who often have fear or aversion against the subject. The computer takes away the learning environment from competition and provides neutral feedback. Dybuster uses external motivators (collection of points) to keep the children motivated. Calcularis relies on intrinsic motivation gained through learning progress.

3. Children with learning disabilities usually have heterogeneous performance profiles, thus a high grade of individualization is necessary. Both Calcularis and Dybuster include student models enabling adaptation to the knowledge level of the user.

The three principles are not only important for training programs for children with learning disabilities, but for learning programs in general. As described above Dybuster and Calcularis share the same teaching principles, but implement them differently. One additional common feature of the two programs is their multi-modal approach.

MODELLING ENGAGEMENT DYNAMICS IN SPELLING LEARNING

For the Dybuster learning environment, we have already developed a model for engagement dynamics (Baschera et al., 2011). The model relates input behaviour to learning and explains the dynamics of engagement states. By quantitatively relating input behaviour and learning, the model enables a prediction of focused and receptive states as well as forgetting.

Approach

The approach used for modelling the engagement is articulated in four steps: (1) description of training process; (2) specification of extracted features; (3) feature processing based on domain knowledge; (4) feature selection and model building.

Experimental data

The analysis is based on the input data of a large-scale study in 2006 (Kast et al., 2007). The log files span a time interval of several months, which permits the analysis of multiple time scales: from seconds to months. The German-speaking participants, aged 9-to-11, trained for a period of three months, with a frequency of four times a week, during sessions of 15-to-20 minutes. On average, each user performed approximately 950 minutes of interactive training. The training predominantly took place at home, except once per week, when the children attended a supervised session at our laboratory to ensure the correct use of the system. Due to technical challenges, a subset of 54 log files were completely and correctly recorded (28 dyslexic and 26 control). This dataset records $159\,699$ entered words, together with inputs, errors, and respective timestamps.

Table 1

Extracted features and abbreviations (bold) used in the following.

Feature	**Description**
Timing	
Input **R**ate	Number of keystrokes per second.
Input **R**ate **V**ariance	Variance of the **IR**.
Think **T**ime	Time from dictation of word to first input letter of student.
Time **f**or **E**rror	Time from last correct input letter to erroneous input letter.
Time **t**o **N**otice **E**rror	Time from error input letter to first corrective action.
Off **T**ime	Longest time period between two subsequent letter inputs.
Input & Error Behaviour	
Help **C**alls	Number of help calls (repeating the dictation).
Finished **C**orrectly	True if all errors are corrected when enter key is pressed.
Same **P**osition **E**rror	True if multiple errors occur at one letter position of a word.
Repetition **E**rror	State of previous input of the same word (three states: *Correct / Erroneous / Not Observed).*
Error **F**requency	Relative entropy (Kullback and Leibler, 1951) from observed to expected error distribution (given by the student model (Baschera and Gross, 2010)) over last five inputs. Positive values are obtained from larger errors numbers, negative values from smaller ones.
Controller Induced	
Time **t**o **R**epetition	Time from erroneous input to respective word repetition.
Letters **t**o **R**epetition	Number of entered letters from erroneous input to respective word repetition.

Feature extraction

We identified a set of recorded features which are consistent with previous work (Baker et al., 2004; Johns and Woolf, 2006; Arroyo and Woolf, 2005). Table 1 lists the features, which are evaluated for each word entered by the learner. The set contains measures of input and error behaviour, timing, and variations of the learning setting induced by the system controller which influence the engagement states. While very fast typing can indicate a lack in concentration, very slow typing could also relate to unfocusedness (**IR** and **IRV**). If **TT** is large, the student might not actually be thinking, but just unfocused. The same reasoning holds for the other timing features (**TfE**,**TtNE**,**OT**). Boredom or a lack of motivation can lead to many help calls (**HC**). And a lack of concentration usually results in more errors (**FC**, **SPE**, **RE**, **EF**). Finally, the controller induced features are important, as they indicate the time between repetitions and thus have a direct influence on forgetting.

Engagement states are inferred from the repetition behaviour of committed errors and without external direct assessments. We subscribe to the validated hypothesis of interplay between human learning and affective dynamics (Kort et al., 2001). Committed errors and the knowledge state at subsequent spelling requests of the same word are jointly analysed. Error repetition acts as a noisy indicator for learning and forgetting. We restrict the analysis to phoneme-grapheme matching (PGM) errors (Baschera and Gross, 2010), which is an error category representing missing knowledge in spelling, in contrast to, e.g., typos. We extracted $14\,892$ observations of PGM errors with recorded word repetitions from the log files.

Feature processing

The processing of continuous features is based upon the following central assumptions: emotional and motivational states come in spurts (Johns and Woolf, 2006), and they affect the observed features on a short-to-medium time scale. Thus, a time scale separation is performed: We distinguish between sustainable progress in the observed input behaviour ($f(i)$) and other local effects ($p(x_i)$), such as the influence of engagement states. The two effects combine linearly to

$$t(x_i) = f(i) + p(x_i), \tag{1}$$

with independent additive normal $p(x_i) \sim \mathcal{N}(0, \sigma^2)$. The transformation $t(\cdot)$ of the original feature x_i consists of scaling and outlier detection. The separation of long-term variation $f(i)$ depends on the temporal input position i in the student input history. The finally obtained additive terms $p(x_i)$ are referred to as processed features. Table 2 lists the employed processor modules. The logarithmic (log) and exponential (exp) transforms reduce the differences in extreme values. The logarithmic transform is for example used for the **IR** feature, as this feature shows a high variance. The splitting operation ($\mathrm{I}_{x>s}$) is applied on the **HC** feature to make it binary (zero/non-zero). Outlier detection is performed to remove extremely large values (if the student leaves the computer, **OT** will be very large). The regression subtraction serves for removing long-term training effects: One important observation here is that children increase their typing velocity over the course of the training, an effect, which is removed through the regression subtraction. A curve is fit to the data using exponential regression and in a second step subtracted from the data. The low-pass and variance filters finally enable a separation of low frequency components from rapid fluctuations of the processed features. A low pass filter is applied to the **TfE** feature to remove short-time effects (high frequency components of the feature).
The selection of processing steps and corresponding coefficients for each feature are the result of a downhill simplex optimization of the differential entropy (with fixed variance) (Nelder and Mead, 1965; Bishop, 2006), resulting in a distribution of $p(x_i)$ with maximal normality. Figure 4 illustrates the processing of the Time for Error (TfE) feature.

Feature selection and model building

The relation between processed features $p(x_i)$ and error repetition γ_r is estimated via LASSO logistic regression (Bishop, 2006) with 10-fold cross-validation for different filter and filter parameters. The regression parameters are denoted by b_i. Figure 5 illustrates the comparison between Error Repetition Probability (ERP) predictions obtained from unprocessed and processed features. The model based on processed features exhibits a better BIC score ($-6\,369$) compared to unprocessed regression ($-6\,742$). In the selected features (see Tab. 3), we identified three main effects influencing the knowledge state at the next repetition:

Focused state indicates focused or distracted state of the student. In a non-focused state more minor errors due to lapse of concentration occur, which are less likely to be committed again at the next repetition (lower ERP).

Receptive state indicates the receptiveness of the student (receptive state or beyond attention span). Non-receptive state inhibits learning and causes a higher ERP.

Table 2

Employed feature processing modules and abbreviations (bold). The parameters of the different operations are learnt from the data. $\mathrm{I}_{x>s}$ denotes the indicator function, var stands for the variance.

Module	**Operation on feature x**	**Parameters**
Scaling		
Logarithmic	$\log(s+x)$	s
Exponential	$\exp(-\frac{a+x}{b})$	a, b
Splitting	$\mathrm{I}_{x>s}$	s
Outlier detection		
Deviation **C**ut	$\min(\mu+\sigma, \max(\mu-\sigma, x)) \quad \mu = \mathrm{mean}(x)$	σ
Regression subtraction		
Learning **C**urve	$x_i - f(i) \quad f(i) = a\exp(-bi) + c$	a, b, c
Filtering		
Low-**P**ass	$x_i = \sum_{j=0}^{n} x_{i-j} G(j,n)$ [1]	n
Variance	$x_i = \mathrm{var}([x_{i-n}, ..., x_i])$	n

[1] $G(j, n)$ corresponds to the sampled Gaussian kernel $G(j,n) = \frac{1}{\sqrt{2\pi n}} e^{-\frac{j^2}{2n}}$.

Table 3

Optimal processing pipeline (applied processing modules ordered from left to right), estimated parameter *b* and significance for features selected by the LASSO logistic regression. Note that the exponential scaling inverts the orientation of a feature. The last two columns show the influence of the engagement states on the features modelled in the DBN: for binary nodes the probability p_1 of being *true*; for Gaussian nodes the estimated mean m of the distribution.

Feature	**Processing Pipeline**	**b**	**sig.**	$\mathbf{p_1[\%]/m}$	
Focused State				focused	non-f.
EF	Exp	0.06	2e-4	0.16	-0.34
IR	Log - DevC - LearnC - Var	-0.12	4e-6	-0.41	0.87
IRV	Log - DevC - LearnC	-0.22	2e-11	-0.36	0.78
REc		-0.28	8e-8	45%	32%
TfE	Log - DevC - LearnC - LowP	-0.50	1e-9	-0.13	0.28
Receptive State				receptive	non-r.
FC		-0.49	1e-7	95%	88%
HC	Split(zero/non-zero)	0.29	2e-4	4%	28%
OT	Log - DevC - LearnC - LowP	0.27	1e-9	-0.35	1.20
REe	LowP	0.20	1e-9	0.07	-0.24
TtNE	Exp - DevC - LearnC	-0.18	1e-5	0.11	-0.36
Forgetting					
TtR	Exp	-0.29	2e-8		
LtR	Log	0.34	1e-9		

Forgetting: the time (decay) and number of inputs (interference) between error and repetition induce forgetting of learned spelling and increase the ERP.

The parameters of the logistic regression indicate how features are related to the ERP. We inferred the affiliation of features to engagement states based on the relations extracted from the regression analysis and expert knowledge about desired input behaviour. For example, the parameter $b = 0.06$ of **EF** demonstrates that a higher than expected error frequency is related to a lower ERP, which indicates that a

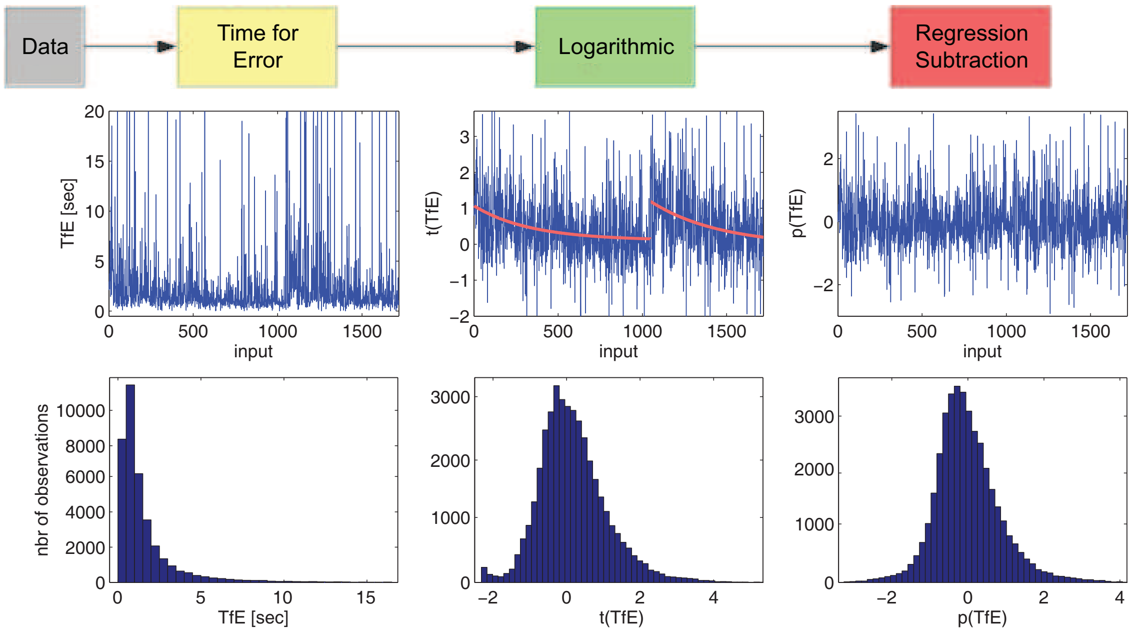


Fig.4. The top line exemplifies the processing pipeline for the **TfE** feature. In the second row, the signal of the processing steps is plotted for the data recorded from two learners: The left most image displays the extracted features, the center image shows the signal after the logarithmic transformation. The red lines denote the fitted exponential regressions for the two learners. The signal after having subtracted the regression fit (denoted by the red lines) is displayed on the right. The third row shows the respective signals for the data of all 54 students: Histogram of extracted features (left), histogram of features after logarithmic transformation (center) and the final histogram after performing the regression subtraction (right). After having processed the feature, the distribution can be well approximated by a normal distribution.

student is non-focused and commits more but rather non-serious errors. By contrast, if a student does not finish an input correctly ($\mathbf{FC} = 0$), the ERP increases ($b = -0.49$). This relation indicates that students are less likely to pick up the correct spelling, when they are not correcting their spelling errors.

In the following we investigate the mutual dependence of the two engagement states, which are considered as nodes in a dynamic Bayesian network (DBN). We compared three models: (1) based on a mutual independence assumption ($\mathrm{F} \leftrightarrow \mathrm{R}$); (2) with dependence of focused state on receptivity ($\mathrm{F} \leftarrow \mathrm{R}$); (3) with dependence of receptivity on focused state ($\mathrm{F} \rightarrow \mathrm{R}$). The parameters of the DBN are estimated based on the expectation maximization (EM) algorithm implemented in Murphy's Bayes net toolbox (Murphy, 2001). The mutual dependence of the engagement states is inferred based on the estimated model evidence (BIC).

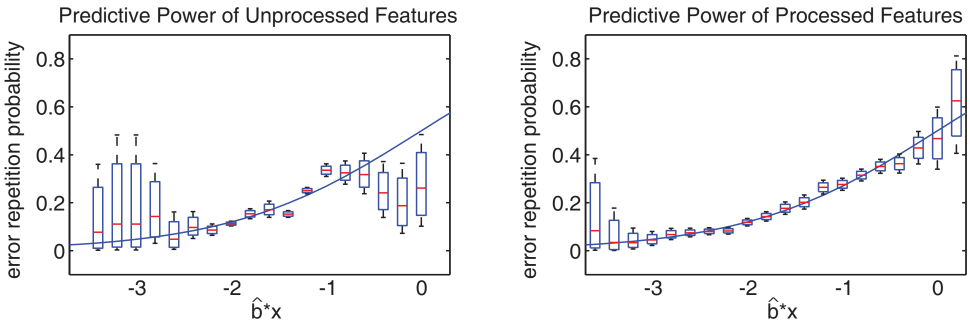


Fig.5. ERP prediction (10-fold cross-validation) from unprocessed (left) and processed features (right). Predictions are plotted as blue curve and accompanied by mean (red stroke), 68% (box), and 95% confidence intervals (whisker) of the observed repetitions for bins containing at least 10 observations. The x-axis is denoted by $\hat{b} * x$, where $\hat{b}$ are the parameters fitted in the LASSO logistic regression and x is the vector of features.

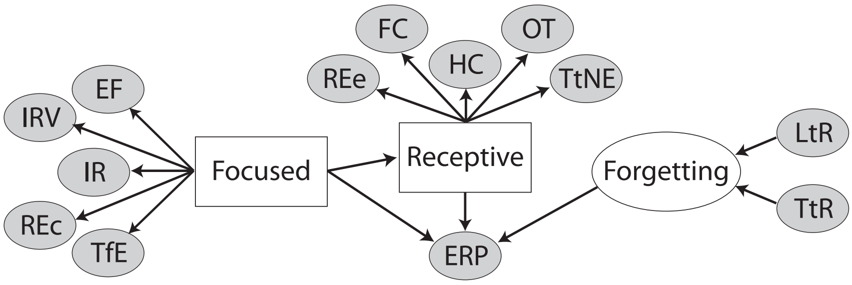


Fig.6. The selected dynamic Bayesian net representation. Rectangle nodes denote dynamic states. Shaded nodes represent observations.

Results of the engagement dynamics in spelling

Figure 6 presents the graphical model (F $\rightarrow$ R) best representing the data with a BIC of $-718\,577$, compared to $-724\,111$ (F $\leftrightarrow$ R) and $-718\,654$ (F $\leftarrow$ R). The relation between the Focused and Receptive state is illustrated by their joint probability distribution in Fig. 7 (left). In a fully focused state, students are never found completely non-receptive. In contrast, students can be distracted (non-focused) despite being in a receptive state.

The ERP conditioned on the two states is presented in Fig. 7 (right). One can observe that the offset between top plane (forgetting) and bottom plane (no forgetting) is greater in the focused compared to the non-focused state. This result underpins the assumption that more non-serious errors are committed in the non-focused state, despite the fact that the correct spelling is actually already known by the student. Therefore, the forgetting has a lower impact on their ERP. As expected, the non-receptive state generally causes a higher ERP. Again, this effect on learning is reduced for non-serious errors in the non-focused state. The estimated parameters of the conditional probability distributions for all the other observed nodes are presented in Tab. 3 (right).

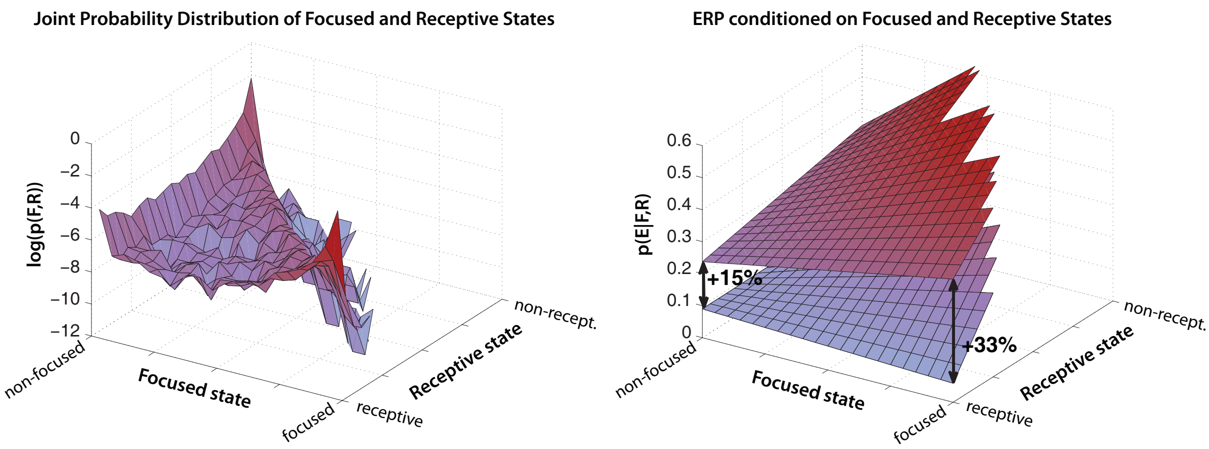


Fig.7. Left: joint probability distribution of Focused and Receptive states. Right: ERP conditioned on engagement states for forgetting (top) and no forgetting (bottom plane). The ERP is plotted for all observed combinations of engagement states only.

The investigation of the age-dependence of engagement states shows that students below the median of 10.34 years exhibit a significantly ($p < 0.001$) higher probability of being classified as non-receptive (24.2%) and non-focused (32.5%) compared to those above the median (20.0% and 27.0%, respectively). This analysis indicates that younger students tend to fall into non-focused and non-receptive states significantly more frequently.

GENERAL ENGAGEMENT DYNAMICS MODELLING FRAMEWORK

To define a framework for building general engagement dynamics model, we extract and analyse the main steps of the introduced model for engagement dynamics in spelling learning. In brief, we can define the following framework:

1. **Indicator definition**: An indicator variable, giving an indication of the engagement state of the children needs to be determined to label the data. This variable can be measured using sensor data (Cooper et al., 2010; Heraz and Frasson, 2009) or by relying entirely on input data as in our engagement model for spelling learning. Entirely data-driven indicators are usually noisy and highly dependent on the learning domain.

2. **Feature extraction**: A set of recorded features needs to be extracted. This set contains measures of input and error behaviour, timing, and variations of the learning setting induced by the system controller. Possible features were proposed in previous work (Baker et al., 2004; Johns and Woolf, 2006; Arroyo and Woolf, 2005; Baschera et al., 2011). The set of meaningful features is strongly influenced by the learning environment.

3. **Feature selection**: To select the features, the relation between the extracted features and the indicator variable needs to be estimated, for example by using a LASSO logistic regression.

4. **Model building**: In a final step, the graphical model needs to be inferred from data. The parameters of the DBN can be estimated using expectation maximization. The quality of different graphical models can be assessed by their BIC score. Model validation can also be performed with Approximation Set Coding (Haghir Chehreghani et al., 2012).

This framework gives an overview of the steps to be taken in order to build a model for engagement dynamics in any domain. Steps 1 and 2 are essential when trying to find a valid model. These two initial steps, however, are also highly dependent on the particular learning domain and the learning environment. The indicator function and the set of features that we applied for the engagement model in spelling are not directly applicable to other domains (such as learning mathematics) for the purpose of modelling engagement dynamics.

ENGAGEMENT MODEL FOR MATHEMATICS LEARNING

Constructing a model of engagement dynamics requires a generic framework to support generalization of engagement behaviour. We start by referring to the previously developed model for engagement dynamics in spelling learning and explore its re-usability. As discussed above, steps 1 and 2 of the general framework are essential. They highlight the dependence on the learning domain and on the specific environment. Thus, a careful comparison between the learning domains, the student models, as well as the available experimental data has to be conducted to decide which parts of the existing model for spelling learning can be reused. Furthermore, we assess the limitations of the existing model and provide suggestions on how to overcome them.

Learning domain

Spelling learning

Spelling a word can be seen as translating from spoken language to written language. In an alphabetic language, like for example English or German, the spoken phonemes need to be matched to graphemes. This matching is not unique because some phonemes can be matched to several graphemes (for instance, the phoneme /f/ can be matched to the graphemes ‘f’ and ‘v’ in German). For spelling learning, different models have been proposed so far. One model for instance suggests that spelling is learnt through the identification of implicit and explicit rules (Hilte and Reitsma, 2011; Ehri, 2000; Cassar and Treiman, 1997; Landerl and Reitsma, 2005; Pacton et al., 2001). Children build up a mental print lexicon, but also abstract regularities from print and are taught rules that underlie their spelling system. It has been shown that children already use phonological and morphological rules from an early age. Another model suggests that spelling is learnt by analogy (Bosse et al., 2003; Campbell, 1985; Marsh et al., 1980; Martinet et al., 2004; Nation and Hulme, 1996, 1998). In this model, the spelling of new words is learnt by analogy to known words called reference words. Both of these presented models imply that spelling learning is a rather *non-hierarchical* process. Rather than learning and understanding concepts and strategies that build up on each other, the process consists of memorizing the phoneme-grapheme matching and its irregularities or of building analogies to existing words.

Mathematics learning

Current neuropsychological models postulate the existence of distinct representational modules. These modules are located in different areas of the brain and are relevant for adult cognitive number processing and calculation. They are activated according to the particular needs of given tasks. In this context, a widely known model is the 'triple-code model' (Dehaene and Cohen, 1995), which assumes three modules for number processing: a verbal number representation supporting verbal counting and number fact retrieval, a visual-Arabic number representation required for solving written arithmetic, and an analogue magnitude representation (spatially organized abstract number line) for semantic number processing. The three representational modules denote the end-state of the learning process, the 'four-step developmental model' (von Aster and Shalev, 2007) describes the path to this final state. This developmental model suggests that the relevant modules develop hierarchically over time depending on the growing capacity and availability of domain-general functions like attention, working memory and processing speed. Rather, Kucian and Kaufmann (2009) suggest an increasing overlap of the different number representations over time. Mathematics learning is, however, not only hierarchical with respect to number processing. Studies have shown that so called precursor abilities such as counting or subitizing are crucial for later mathematical understanding (Landerl et al., 2004; Mazzocco and Thompson, 2005). Strategies develop over time also in the domain of arithmetic operation: children start with simple counting strategies and proceed to more mature strategies and fact retrieval (Carpenter and Moser, 1984; Beishuizen et al., 1997). To summarize, there exists convincing evidence of the fact that learning of mathematics is *hierarchical* in nature: knowledge builds on top of previously learnt concepts. If a step is missing, later steps cannot be learnt effectively.

Student model

Dybuster

In Dybuster, the selection of words to be prompted is adapted to the skill level of the children. The word selected to be trained next is the word with the highest progress potential with respect to training time. The knowledge representation is an estimate of individual mal-rule difficulties. Mal-rules define different error types which a child can commit. Possible error categories are, e.g., capitalization errors, typing errors (depending on key distance or for technical reasons), letter confusion (visual or auditory similarity) or erroneous phoneme-grapheme matching. As immediate feedback is presented after an erroneous letter, error classification is ambiguous, i.e., different deficits can lead to the same final error. To deal with this ambiguity, Dybuster uses an inference algorithm for perturbation models based on Poisson regression (Baschera and Gross, 2010). The algorithm is designed to handle unclassified input with multiple errors described by independent mal-rules. During the training, the representation of the student's mastery of the domain is continuously updated after each entered word. Based on these estimates, a prediction of further spelling performance and a classification of committed errors for each individual student can be estimated. In addition to this spelling knowledge representation, the word selection controller accounts for the optimal time to repetition (time until a previously misspelled word is repeated).

Calcularis

In Calcularis, the selection of games and tasks is adapted to the skill level of the child. The mathematical knowledge of the user is modelled using a DBN. This network consists of a directed acyclic graph representing different mathematical skills and their mutual dependencies. The resulting student model contains 100 different skills. Each skill can have two states: a learnt state and an unlearnt one. The probabilities for these states are inferred by posing tasks and evaluating user actions. After each solved task, the system updates the posterior probabilities of the skills. Describing the structure of the student model as a graph suggests that the selection of actions is rule-based and non-linear. Based on the current state, three possible actions can be selected: going back to an easier skill, going forward to a more difficult skill, or additional training of the current skill. This decision is based on lower and upper thresholds. If the option of ‘go back’ or ‘go forward’ is selected, the system selects the next candidate skill on the basis of the built-in system of rules. It has been shown that this type of control mechanism is beneficial for the children’s progress (Käser et al., 2012). An additional feature of the student model is the attached bug library which is directly integrated with the DBN. The system is able to recognize typical errors that children commit in arithmetic.

Experimental data

The available experimental data depend on the learning environment and the student model used, and thus indirectly also on the learning domain and the properties of the respective learning disability. For both learning environments, log file data have been collected during evaluation. The experimental data for Dybuster comprise 54 log files (54 participants), recording approximately 950 minutes of interactive training per user and consisting of $159\,699$ entered words, together with inputs, errors, and respective timestamps. For Calcularis, 96 log files (96 participants) containing approximately 600 minutes of computer-based training per user, giving a total of $144\,000$ tasks, are available. Calcularis records all tasks together with inputs, errors and respective timestamps. Furthermore, all user inputs (including careless keystrokes and mouse clicks) are recorded.

Although the experimental data looks similar for both learning environments at first glance, it is quite different when extracting possible features. For the Dybuster learning environment, features such as the typing velocity or the answer time can easily be compared across words. For the Calcularis environment, the situation is different. Calcularis uses a hierarchical skill model. Tasks associated with different skills have different properties and difficulty levels and thus features cannot be easily compared over different task types. The answer time for an addition task, such as ‘3+4’, will always be shorter than for the addition ‘53+39’. This fact also limits the number of available samples for comparison as the samples need to be divided onto the different skills. Moreover, the Calcularis learning environment features games with different input possibilities such as a joystick, mouse click or keyboard input. This increases the variability of the data but in turn also makes comparison between different tasks more challenging.

Engagement model - limitations and extensions

The comparisons conducted in the previous sections have shown that there are significant differences in the two learning domains, the modelling of the student in the environment, as well as in the availability and interpretation of the experimental data. Given this information, we will, in the following, assess the first two steps of the engagement model for spelling learning to identify the parts of the model that can be reused. Furthermore, we define desirable properties of an indicator function (step 1) and a feature set (step 2) applicable to learning in general and make a first draft of a possible general feature set.

Indicator function

The model for engagement dynamics in spelling learning uses the error repetition probability (ERP) as a noisy indicator. If the student is in a distracted state, more careless errors will occur which are unlikely to be repeated (low ERP). If the student is in a non-receptive state (inhibits learning), committed errors will probably be repeated (high ERP). This indicator function is meaningful under the following (strict) assumptions:

- Stationary learning environment: The learning environment consists of only one type of task (here the typing of words).
- Non-hierarchical learning domain: The learning works in a non-hierarchical way, for example through memorization. This assumption means that a word is learnt through memorizing the spelling in the case of Dybuster.

The learning environment for mathematics learning as well as the learning domain do not fulfill these properties. Calcularis consists of a number of skills at varying difficulty levels, each of them depending on each other. Performance or error measures can thus not easily be compared across the different skills. Furthermore, mathematics learning is very hierarchical. Besides knowing rules or building procedural knowledge, conceptual knowledge (understanding the ‘why’) needs to be built. If a child makes an error such as ‘12-5=3’, it makes no sense to repeat exactly the same task after a certain amount of time. The child needs to learn how a ten-crossing is handled. Having learnt that, the child can solve all tasks involving a ten-crossing.
How should an appropriate indicator function look for a hierarchical learning domain and a learning environment employing different skills? Why do we need an indicator function in the first place? As we rely on input data only, no ground truth about the emotional state of the user is available. The indicator function represents the emotional state (for example engagement) over time and thus provides us with a labelling of the data and therefore we deal with supervised learning instead of unsupervised learning. Assuming an interplay between human learning and affective dynamics (Kort et al., 2001), an indicator based on performance measures in the learning environment can be selected. However, being in an engaged state is a necessary but not sufficient condition for learning. The indicator function therefore

needs to differentiate between different reasons for low progress in the environment. Besides not being engaged, the tasks posed can be too easy or too difficult (not matching the skill level of the user) or there can be task comprehension problems. All these cases need to be taken into account. Furthermore, the indicator function needs to consider the hierarchical structure of the skills and the dependencies among them and thus also account for previous knowledge. Still, an indicator function relying purely on input data will always be an approximation of ground-truth. The input data can, however, be enhanced to increase the reliability of the indicator. Calcularis, for example, also records careless input of the children such as random key strokes or mouse clicks. These inputs give an additional indication of the engagement state.

Feature set

The set of features used for the engagement dynamics model in spelling learning is very specific and in particular also very much adapted to the learning domain and the learning environment used. The features used can be divided into three categories. Features in the *Timing* category are useful to indicate attention, but also particularly specific to the learning environment. Features such as the input rate and its variance assume an environment where the results are entered via the keyboard and where the typing velocity is meaningful, which is not the case for the mathematics learning environment. Also features such as **TfE** and **TtNE** assume an immediate feedback on the error (before the child has typed the whole result). On the other hand, think time and off time indicate the child's performance also in the mathematics learning environment. Also in the second category focusing on *Input & error behaviour*, only few features can be re-used. Help calls are for example not possible in every environment. The **FC** feature is only meaningful if feedback on errors is given already while the child enters the result. And the **SPE** feature is specific to the learning domain. In contrast, features such as repetition error or error frequency describe general error behaviour (Does the user repeat errors? How many errors does the user make?) and thus are meaningful for any learning domain. The third category (*Controller Induced*) is completely dependent on the learning environment, as these features are induced by the controller of the particular environment. Table 4 discusses which features are specific to the learning domain and the environment of spelling learning, and which features could be reused for the mathematics learning environment.

As is evident from the table, the given feature set is specifically designed for the spelling learning environment, yielding very good results. For this reason, most of the features cannot be directly applied to a different learning domain or a different learning environment such as mathematics learning. However, we can divide the features designed for the spelling learning environment into different feature categories and derive a general feature set from those. We use the categories *input behaviour, problem statement, problem-solving behaviour, performance* and *environment*. Table 5 shows the categories as well as our suggestion for a general feature set associated with these categories for engagement dynamics modelling.

The features of the comprehensive feature set can be used for different learning domains and environments and are particularly suitable for hierarchical learning domains such as mathematics learning. Table 6 shows that most features could be directly applied to the mathematics learning environment, such as the one provided by Calcularis.

Table 4
Assessment of feature set for the engagement dynamics model in spelling learning.

Feature	**Assessment**	**Reason**
Timing		
Input **R**ate	No	Input rate not meaningful for mathematics learning.
Input **R**ate **V**ariance	No	Same reason as for the **IR**.
Think **T**ime	Yes	Can be replaced by answer time, i.e., the time the child needs to answer the task.
Time **f**or **E**rror	No	Only meaningful in an environment with immediate feedback on errors.
Time **t**o **N**otice **E**rror	No	Feedback is only given after the whole result has been entered.
Off **T**ime	Yes	Could be redefined to be the time until the child starts answering the task.
Input & error behaviour		
Help **C**alls	No	No help calls possible in the environment.
Finished **C**orrectly	No	Feedback is only given after the whole result has been entered.
Same **P**osition **E**rror	No	Only meaningful for spelling learning.
Repetition **E**rror	Yes	Might be replaced by assessing the previous opportunity the child had to apply a certain skill.
Error **F**requency	Yes	Student model needs to compute expected error distribution.
Controller Induced		
Time **t**o **R**epetition	No	Repetition of exactly same task is not done.
Letters **t**o **R**epetition	No	Repetition of exactly same task is not done.

DISCUSSION

In this paper, we introduced a framework for modelling engagement dynamics in spelling learning. We discussed possible extensions in the context of learning in mathematics. The study explores the idea of transferring existing results in the context of engagement modelling to general applications for learning disabilities. Our assumptions are scientifically justified by the significant co-occurrence of dyslexia and dyscalculia with ADHD and the similar implications such as anxiety and low intrinsic motivation of the two learning disabilities. This observation constitutes a clear indicator of the existence of similar engagement dynamics, thereby suggesting general measures and models of engagement.

We performed a detailed analysis of similarities and differences of the two disabilities. In the following, we present a summary of our findings and finish with a short conclusion.

Summary

In this paper, we argued that similar engagement patterns can be assumed for developmental dyslexia and developmental dyscalculia. On the basis of the available justifications, it follows that a similar engagement model for both learning disabilities would be favourable. Our analysis of the learning domain and the learning environments, of their corresponding student models, as well as of the experimental data, suggests that the proposed framework is suitable for the case of developmental dyscalculia. Our findings show, however, that indicator functions and features are specific to the learning domain. Table 7 summarizes the similarities and dissimilarities of the two cases.

From this comparison we conclude that there are substantial differences in the learning domain, which in turn directly influence the learning environment and the student model. Furthermore, these

Table 5
Sketch of a general feature set (abbreviations in bold) for engagement dynamics modelling.

Generalized feature	**Description**
Input behaviour	
Input **T**ype	The type of the input, e.g., mouse, keyboard, pull-down menu, etc.
Valid **I**nput	True if the input is valid, e.g., input string only contains numbers.
Input **S**tatistics	Statistics of the input as for example mean input rate or input rate variance.
Problem-oriented **I**nput	True if the input is related to the problem, e.g., user enters text into the answer.
Problem statement	
Problem **D**ifficulty	Ideally an overall measure of the problem difficulty.
Problem **T**ype	The kind of problem at hand.
Problem **F**amiliarity	True if the user is familiar with the kind of problem.
Problem-solving behaviour	
Time to **S**olution	Total time spent on this problem until solution.
Time **L**ast **S**olutions	Total time spent on the last n problems.
Time **D**eviation	Standard deviation from mean time to solution for this kind of problem.
Answer **T**ime	Time until user starts answering the problem after she sees the problem statement.
Problem **A**pproach	The user's approach to the problem, e.g., trial and error, systematic, etc.
Help **U**sage	If a help system is available how is it used, e.g., frequency of use.
Performance	
Correctness of **A**nswer	Assessment of user answer: correct, wrong or misconception.
Answer **A**ssessment	User performance meets model expectations (e.g., posterior probability).
Error **I**nformation	Information about the committed error, e.g., spelling error.
Error **R**epetition	Number of errors in the past for the same kind of problem.
Error **F**requency	Frequency of certain error types.
Error **C**ount	Number of errors that are similar to the current error in the last n problems.
Environment	
Time **B**etween **P**roblems	Time from last similar problem to this one.
Similar **P**roblems **C**ount	Number of problems that were similar to the current one in the last n problems.
Work **B**etween **P**roblems	Amount of work between the current and the last similar problem.
Session **D**uration	Duration of the training session.
Time of the **Day**	Time of the day the training session takes place.

differences indirectly affect the experimental data as well. Therefore, the application of the indicator function and of the feature set specified for the model of engagement dynamics in spelling learning is fairly sophisticated. Rather, a more general indicator function and a comprehensive feature set need to be defined. At present, this is an area of active research.

Conclusion

This study presented a model for engagement dynamics for spelling learning and its extension. We raised the question of how a general framework for modelling engagement dynamics in learning could be defined, focusing on developmental dyslexia and developmental dyscalculia. Our results emphasize that

Table 6

Assessment of the general feature set for the mathematics learning environments.

Generalized feature	Assessment	Reason
Input behaviour		
Input **T**ype	Yes	Mouse, keyboard, joystick.
Valid **I**nput	Yes	Input is a valid number.
Input **S**tatistics	No	Input statistics not meaningful for the specific environment.
Problem-oriented **I**nput	Yes	Click at the right place (where you should click).
Problem statement		
Problem **D**ifficulty	Yes	Can directly be derived from the student model.
Problem **T**ype	Yes	Trained skill.
Problem **F**amiliarity	Yes	True if the user has trained the same skill before.
Problem-solving		
Time to **S**olution	Yes	Directly applicable.
Time **L**ast **S**olutions	Yes	Directly applicable.
Time **D**eviation	Yes	Directly applicable.
Answer **T**ime	Yes	Directly applicable.
Problem **A**pproach	Yes	Problem omission can be detected.
Help **U**sage	No	No help system available.
Performance		
Correctness of **A**nswer	Yes	Directly applicable.
Answer **A**ssessment	Yes	Comparison of student's performance against estimated model performance.
Error **I**nformation	Yes	Directly applicable using the bug library.
Error **R**epetition	No	Repetition of exactly same task is not done.
Error **F**requency	Yes	Directly applicable using the bug library.
Error **C**ount	Yes	Directly applicable using the bug library.
Environment		
Time **B**etween **P**roblems	Yes	Time from last problem that trained the same skill to this one.
Similar **P**roblems **C**ount	Yes	Number of problems that trained the same skill in the last n problems.
Work **B**etween **P**roblems	Yes	Amount of work between last problem that trained the same skill and this one.
Session **D**uration	Yes	Directly applicable.
Time of the **Day**	Yes	Directly applicable.

the indicator function and the feature extraction are particularly important for selecting a valid model. A closer comparison highlighted, however, the strong dependency of these two steps on the learning domain, the student model that is used, and the available experimental data. The conducted comparison illustrates that there are significant differences in the learning environments which prevent a straightforward application of the engagement model in spelling learning onto mathematics learning. We defined desirable properties of a general indicator function and proposed a comprehensive feature set in order to exploit the increased flexibility that is provided by such a general engagement model.

Table 7
Comparison of the two cases of developmental dyscalculia and dyslexia.

Category	Dyslexia	Dyscalculia
Learning disability	Brain-based disorder Comorbidities (Dyscalculia, ADHD) Aversion & anxiety against the subject	Brain-based disorder Comorbidities (Dyslexia, ADHD) Aversion & anxiety against the subject
Learning domain	Static (non-hierarchical) Learning through memorization & analogies	Hierarchical Conceptual knowledge important
Learning environment	One main learning game Multi-modal cues recode textual input string Difficulty of word adapted to user	Range of games ordered hierarchically Visual cues encode properties of number Selection of games and tasks adapted to user
Student model	Poisson-based perturbation model Selection of word with highest progress potential	Dynamic Bayesian network Non-linear, rule-based task selection
Experimental data	Input logs with inputs, errors and timestamps Input from keyboard No additional information	Input logs with inputs, errors and timestamps Input from keyboard, mouse and joystick Recording of invalid inputs

ACKNOWLEDGEMENTS

The work presented in this paper was funded by the CTI-grants 8970.1 and 11006.1, as well as with grants from the Swiss SystemsX.ch initiative.

REFERENCES

Ackerman, P. T. and Dykman, R. A. (1995). Reading-disabled students with and without comorbid arithmetic disability. *Developmental Neuropsychology*, 11:351–371.

Arroyo, I. and Woolf, B. P. (2005). Inferring learning and attitudes from a Bayesian Network of log file data. In *Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*, pages 33–40.

Bacon, A. M., Handley, S. J., and McDonald, E. L. (2007). Reasoning and dyslexia: a spatial strategy may impede reasoning with visually rich information. *British Journal of Psychology*, 98:79–92.

Badian, N. A. (1983). Arithmetic and nonverbal learning. In *Progress in Learning Disabilities*, pages 235–264. Myklebust, H. R.

Badian, N. A. (1999). Persistent arithmetic, reading, or arithmetic and reading disability. *Annals of Dyslexia*, 49:45–70.

Baker, R. S., Corbett, A. T., and Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, pages 531–540. Springer Verlag.

Barbaresi, M. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., and Jacobsen, S. J. (2005). Math learning disorder: Incidence in a population-based birth cohort 1976-1982. *Ambulatory Pediatrics*, 5:281–289.

Baschera, G. M., Busetto, A. G., Klingler, S., Buhmann, J., and Gross, M. (2011). Modeling Engagement Dynamics in Spelling Learning. In Biswas, G., Bull, S., Kay, J., and Mitrovic, A., editors, *Artificial Intelligence in Education*, volume 6738 of *Lecture Notes in Computer Science*, pages 31–38. Springer Berlin / Heidelberg.

Baschera, G. M. and Gross, M. (2010). Poisson-Based Inference for Perturbation Models in Adaptive Spelling Training. *International Journal of AIED*, 20(4):333–360.

Beishuizen, M., van Putten, C. M., and van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to one hundred. *Learning and Instruction*, 7(1):87–106.

Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, 25(3-4):399–410.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bosse, M. L., Valdois, S., and Tainturier, M. J. (2003). Analogy without priming in early spelling development. *Reading and Writing*, 16:693–716.

Busetto, A. G., Ong, C., and Buhmann, J. (2009). Optimized Expected Information Gain for Nonlinear Dynamical Systems. *ICML 2009*, pages 97–104.

Butterworth, B. (2005a). The development of arithmetical abilities. *Journal of Child Psycholgy and Psychiatry*, 46:3–18.

Butterworth, B. (2005b). Developmental dyscalculia. In Campell, J. D., editor, *The Handbook of Mathematical Cognition*, pages 455–469.

Bynner, J. (1997). Basic skills in adolescents' occupational preparation. *Career Development Quarterly*, 27:777–786.

Campbell, R. (1985). When children write pseudowords to dictation. *Journal of Experimental Child Psychology*, 57:26–41.

Carpenter, T. and Moser, J. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15(3):179–202.

Cassar, M. and Treiman, R. (1997). The beginnings of orthographic knowledge: children's knowledge of double letters in words. *Journal of Educational Psychology*, 89:631–644.

Cohen Kadosh, R., Cohen Kadosh, K., Schuhmann, T., Kaas, A., Goebel, R., Henik, A., and Sack, A. T. (2007). Virtual dyscalculia induced by parietal-lobe TMs impairs automatic magnitude processing. *Current Biology*, 17:689–693.

Cooper, D., Muldner, K., Arroyo, I., Woolf, B., and Burleson, W. (2010). Ranking Feature Sets for Emotion Models Used in Classroom Based Intelligent Tutoring Systems. In De Bra, P., Kobsa, A., and Chin, D., editors, *User Modeling, Adaptation, and Personalization*, volume 6075 of *Lecture Notes in Computer Science*, pages 135–146. Springer Berlin / Heidelberg.

Dehaene, S. and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, 1:82–120.

Demonet, J. F., Taylor, M. J., and Chaix, Y. (2004). Developmental dyslexia. *Lancet*, 363(9419):1451–1460.

Dirks, E., Spyer, G., van Lieshout, E., and de Sonneville, L. (2008). Prevalence of combined reading and arithmetic disabilities. *Journal of Learning Disabilities*, 41:460–473.

Dowker, A. (2001). Intervention in numeracy: the development of a Numeracy Recovery project for young children with arithmetical difficulties. *Support for Learning*, 16:6–10.

Edwards, L. (2003). Writing instruction in kindergarten: examining an emerging area of research for children with writing and reading difficulties. *Journal of Learning Disabilities*, 36(2):136–148.

Ehri, L. C. (2000). Learning to read and learning to spell: two sides of a coin. *Topics in Language Disorders*, 20:19–36.

Fletcher, J. M. (2005). Predicting math outcomes: Reading predictors and comorbidity. *Journal of Learning Disabilities*, 38:308–312.

Gaab, N., Gabrieli, J. D. E., Deutsch, G. K., Tallal, P., and Temple, E. (2007). Neural correlates of rapid auditory processing are disrupted in children with developmental dyslexia and ameliorated with training: an fMRI study. *Restorative Neurology and Neuroscience*, 25(3-4):295–310.

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., and Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. *Nature Neuroscience*, 9(10):1213–1217.

Galaburda, A. M., Sherman, G. F., Rosen, G. D., Aboitiz, F., and Geschwind, N. (1985). Developmental dyslexia: four consecutive patients with cortical anomalies. *Annals of Neurology*, 18(2):222–233.

Germanò, E., Gagliano, A., and Curatolo, P. (2010). Comorbidity of ADHD and dyslexia. *Developmental Neuropsychology*, 35(5):475–493.

Goswami, U. (2003). Why theories about developmental dyslexia require developmental designs. *Trends in Cognitive Sciences*, 7(12):534–540.

Griffin, S., Case, R., and Siegler, R. (1994). Rightstart: Providing the central conceptual prerequisites for the first formal learning of arithmetic to students at risk for school failure. In McGilly, K., editor, *Classroom lessons: Integrating cognitive theory and classroom practice*. Cabridge, MA: MIT Press.

Gross, M. and Vögeli, C. (2007). A Multimedia Framework for Effective Language Training. *Computer & Graphics*, 31:761–777.

Haghir Chehreghani, M., Busetto, A. G., and Buhmann, J. M. (2012). Information theoretic model validation for spectral clustering. *Journal of Machine Learning Research Proc. 15th AISTATS 12*, pages 495–503.

Hatcher, P. J., Hulme, C., Miles, J. N., Carroll, J. M., Hatcher, J., and Gibbs, S. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: a randomized controlled trial. *Journal of Child Psychology and Psychiatry*, 47(8):820–827.

Heraz, A. and Frasson, C. (2009). Predicting Learner Answers Correctness through Brainwaves Assessment and Emotional Dimensions. In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, pages 49–56.

Hilte, M. and Reitsma, P. (2011). Effects of explicit rules in learning to spell open- and closed-syllable words. *Learning and Instruction*, 21(1):34–45.

Johns, J. and Woolf, B. (2006). A dynamic mixture model to detect student motivation and proficiency. In *Proceedings of the 21st national conference on Artificial intelligence - Volume 1*, AAAI'06, pages 163–168.

Käser, T., Busetto, A. G., Baschera, G.-M., Kohn, J., Kucian, K., von Aster, M., and Gross, M. (2012). Modelling and optimizing the process of learning mathematics. In Cerri, S., Clancey, W., Papadourakis, G., and Panourgia, K., editors, *Intelligent Tutoring Systems*, volume 7315 of *Lecture Notes in Computer Science*, pages 389–398. Springer Berlin / Heidelberg.

Käser, T., Kucian, K., Ringwald, M., Baschera, G.-M., von Aster, M., and Gross, M. (2011). Therapy software for enhancing numerical cognition. In Özyurt, J., Anschütz, A., Bernholt, S., and Lenk, J., editors, *Interdisciplinary Perspectives on Cognition, Education and the Brain*, volume 7 of *Hanse Studies*, pages 219–228. Hanse-Wissenschaftskolleg (HWK),Institute for Advanced Study, Oldenburg: BIS-Publisher.

Kast, M., Meyer, M., Vögeli, C., Gross, M., and Jäncke, L. (2007). Computer-based Multisensory Learning in Children with Developmental Dyslexia. *Restorative Neurology and Neuroscience*, 25(3-4):355–369.

Kaufmann, L., Handl, P., and Thöny, B. (2003). Evaluation of a numeracy intervention program focusing on basic numerical knowledge and conceptual knowledge: A pilot study. *Journal of Learning Disabilities*, 35:564–573.

Kort, B., Reilly, R., and Picard, R. W. (2001). An Affective Model of Interplay Between Emotions and Learning: Reengineering Educational Pedagogy - Building a Learning Companion. *Advanced Learning Technologies*, pages 43–46.

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., Gälli, M., Martin, E., and von Aster, M. (2011). Mental Number Line Training in Children with Developmental Dyscalculia. *NeuroImage*, 57(3):782–795.

Kucian, K. and Kaufmann, L. (2009). A developmental model of number representation. *Behavioral and Brain Sciences*, 32:313–373.

Kucian, K., Loenneker, T., Dietrich, T., Dosch, M., Martin, E., and von Aster, M. (2006). Impaired neural networks for approximate calculation in dyscalculic children: a functional MRI study. *Behavioral and Brain Functions*, 2:31.

Kujala, T., Karma, K., Ceponiene, R., Belitz, S., Turkkila, P., and Tervaniemi, M. (2001). Plastic neural changes and reading improvement caused by audiovisual training in readingimpaired children. *Proceedings of the National Academy of Sciences USA*, 98(18):10509–10514.

Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.

Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8-9-year-old students. *Cognition*, 93:99–125.

Landerl, K. and Reitsma, P. (2005). Phonological and morphological consistency in the acquisition of vowel duration spelling in Dutch and German. *Journal of Experimental Child Psychology*, 92:322–344.

Lehmann, S. and Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Brain Research. Cognitive Brain Research*, 24:326–334.

Lenhard, A., Lenhard, W., Schug, M., and Kowalski, A. (2011). Computerbasierte Mathematikförderung mit den Rechenspielen mit Elfe und Mathis I. Vorstellung und Evaluation eines Computerprogramms für Erst- bis Drittklässler. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43(2):79–88.

Lewis, C., Hitch, G. J., and Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9- to 10-year old boys and girls. *Journal of Child Psychology and Psychiatry*, 35:283–292.

Lorusso, M. L., Facoetti, A., Paganoni, P., Pezzani, M., and Molteni, M. (2006). Effects of visual hemisphere-specific stimulation versus reading-focused training in dyslexic children. *Neuropsychological Rehabilitation*, 16(2):194–212.

Marsh, G., Friedman, M., Welch, V., and Desberg, P. (1980). The development of strategies in spelling. In Frith, U., editor, *Cognitive processes in spelling*, pages 339–354. Academic, London.

Martinet, C., Valdois, S., and Fayol, S. (2004). Lexical knowledge develops from the beginning of literacy acquisition. *Cognition*, 91:11–22.

Mazzocco, M. M. and Thompson, R. E. (2005). Kindergarten Predictors of Math Learning Disability. *Learning Disabilities Research & Practice*, 20(3):142–155.

Murphy, K. (2001). The Bayes Net Toolbox for Matlab. *Computing Science and Statistics*, 33.

Mussolin, C., De Volder, A., Grandin, C., Schlogel, X., Nassogne, M. C., and Noel, M. (2010). Neural correlates of symbolic number comparison in developmental dyscalculia. *Journal of Cognitive Neuroscience*, 22(5):860–874.

Nation, K. and Hulme, C. (1996). The automatic activation of sound-letter knowledge: an alternative interpretation of analogy and priming effects in early spelling development. *Journal of Experimental Child Psychology*, 63:416–435.

Nation, K. and Hulme, C. (1998). The role of analogy in early spelling development. In Hulme, C. and Joshi, R. M., editors, *Reading and spelling: Development and disorders*, pages 433–445. Erlbaum, Mahwah, NJ.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7:308–313.

O'Shaughnessy, T. E. and Swanson, H. L. (2000). A comparison of two reading interventions for children with reading disabilities. *Journal of Learning Disabilities*, 33(3):257–277.

Ostad, S. A. (1997). Developmental differences in addition strategies: A comparison of mathematically disabled and mathematically normal children. *British Journal of Education Psychology*, 67:345–357.

Ostad, S. A. (1998). Comorbidity between mathematics and spelling difficulties. *Logopedics Phoniatrics Vocology*, 23:145–154.

Ostad, S. A. (1999). Developmental progression of subtraction strategies: A comparison of mathematically normal and mathematically disabled children. *European Journal of Special Needs Education*, 14:21–36.

Pacton, S., Perruchet, P., Fayol, M., and Cleeremans, A. (2001). Implicit learning out of the lab: the case of orthographic regularities. *Journal of Experimental Psychology: General*, 130:401–426.

Price, G. R., Holloway, I., Räsänen, P., Vesterinen, M., and Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Current Biology*, 17.

Robichon, F., Besson, M., and Habib, M. (2002). An electrophysiological study of dyslexic and control adults in a sentence reading task. *Biological Psychology*, 59(1):29–53.

Rubinsten, O. and Henik, A. (2005). Automatic activation of internal magnitudes: a study of developmental dyscalculia. *Neuropsychology*, 19:641–648.

Rubinsten, O. and Tannock, R. (2010). Mathematics anxiety in children with developmental dyscalculia. *Behavioral and Brain functions*, 6:46.

Russeler, J., Gerth, I., and Munte, T. F. (2006). Implicit learning is intact in adult developmental dyslexic readers: evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology*, 28(5):808–827.

Santos, A., Joly-Pottuz, B., Moreno, S., Habib, M., and Besson, M. (2007). Behavioural and event-related potentials evidence for pitch discrimination deficits in dyslexic children: improvement after intensive phonic intervention. *Neuropsychologia*, 45(5):1080–1090.

Schulte-Korne, G., Deimel, W., Bartling, J., and Remschmidt, H. (2004). Neurophysiological correlates of word recognition in dyslexia. *Journal of Neural Transmission*, 111(7):971–984.

Shalev, R. and von Aster, M. G. (2008). Identification, classification, and prevalence of developmental dyscalculia. *Enc. of Language and Literacy Development*, pages 1–9.

Shalev, R. S. (2004). Developmental dyscalculia: Review. *Journal of Child Neurology*, 19:765–771.

Shams, L. and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Science*, 12:411–417.

Shaywitz, B. A., Fletcher, J. M., and Shaywitz, S. E. (1994). A Conceptual Framework for Learning Disabilities and Attention-Deficit/Hyperactivity Disorder. *Canadian Journal of Special Education*, 9(3):1–32.

Shaywitz, B. A., Shaywitz, S. E., Blachman, B. A., Pugh, K. R., Fulbright, R. K., and Skudlarski, P. (2004). Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. *Biological Psychiatry*, 55(9):926–933.

Shaywitz, S. E. (1998). Dyslexia. *The New England Journal of Medicine*, 338(5):307–312.

Siegler, R. S. and Ramani, G. B. (2009). Playing linear number board games - but not circular ones - improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, 101:249–274.

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews Neuroscience*, 5(9):721–728.

Uther, M., Kujala, A., Huotilainen, M., Shtyrov, Y., and Naatanen, R. (2006). Training in Morse code enhances involuntary attentional switching to acoustic frequency: Evidence from ERPs. *Brain Research*, 1073-1074:417–424.

Vadasy, P. F., Jenkins, J. R., and Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities. *Journal of Learning Disabilities*, 33(6):579–590.

Van De Rijt, B. A. M. and Van Luit, J. E. H. (1998). Effectiveness of the Additional Early Mathematics program for teaching children early mathematics. *Instructional Science*, 26(5):337–358.

von Aster, M. G. and Shalev, R. (2007). Number development and developmental dyscalculia. *Developmental Medicine and Child Neurology*, 49:868–873.

Wilson, A. J., Dehaene, S., Dubois, O., and Fayol, M. (2009). Effects of an Adaptive Game Intervention on Accessing Number Sense in Low-Socioeconomic-status Kindergarten Children. *Mind, Brain and Education*, 3(4):224–234.

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., and Cohen, D. (2006a). Principles underlying the design of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and*

Brain Functions, 2:19.

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., and Dehaene, S. (2006b). An open trial assessment of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, 2:20.

World Health Organization (1993). *ICD-10: The international classification of diseases.* Geneva: World Health Organization, vol. 10: classification of mental and behavioural disorders edition.

Wright, R. J. (2003). A mathematics recovery: Program of intervention in early number learning. *Australian Journal of Learning Disabilities*, 8(4):6–11.

Ziegler, A., Konig, I. R., Deimel, W., Plume, E., Nothen, M. M., and Propping, P. (2005). Developmental dyslexia-recurrence risk estimates from a german bi-center study using the single proband sib pair design. *Human Heredity*, 59(3):136–143.