SRIP Report

by Dr. Tan Da Yang and Muhammad Danish Syazwan

**Abstract**

The main goal of this research is to investigate and categorize social and emotional learning (SEL) motivations behind problem-solving, based on the metrics introduced in Elliott et al.’s (2018) study on SEL. By using the electroencephalogram (EEG) wearable device and video processing techniques, we can measure the brain activity of the subjects when solving certain problems. We explore the use of the Mind Monitor application to process insightful information from the EEG signals. An experiment was carried out where the subjects solve problems while wearing the EEG headsets. The solving process will also be recorded and processed separately as variables. Machine learning algorithms were then developed to categorize the subjects’ social and emotional motivations. The dataset consists of the EEG readings and processed video recordings and was subsequently split into 80/20 for testing and validation. We hope to use this classification algorithm to predict people’s SEL motivations behind problem-solving.

1. **Introduction**

**1.1 Social and Emotional Learning (SEL)**

It is important for us to examine the social and emotional motivations behind learning and problem-solving to better understand how people react to different situations.

Social and emotional learning (SEL) is defined as the “process of acquiring knowledge, skills, attitudes, and beliefs to identify and manage emotions; to care about others; to make good decisions; to behave ethically and responsibly; to develop positive relationships and to avoid negative behaviors” (Elias & Moceri, 2012). It emphasises a combination of social, emotional, behavioural, character skills, or competencies, to ensure one succeeds in various phases in life such as schools, workplace, and within the community (Frey et al., 2009).

In an effort to address the social and emotional needs of students, Waters and Sroufe (1983) define competence as the ability to “generate and coordinate flexible, adaptive responses to the demands and to generate and capitalize on opportunities in the environment”. This definition includes:

* the individual’s contribution to the situation
* recognition of the opportunity for response
* prior acquisition of response alternatives
* selection from response alternatives
* motivation to respond
* choosing to persist or change the response
* modulation (fine-tuning) or response

Social and emotional skills are “individual characteristics that originate from biological predispositions and environmental factors, are manifested as consistent patterns of thoughts, feelings, and behaviors, continue to develop through formal and informal learning experiences, and that influence important socio-economic outcomes throughout the individual’s life” (De Fruyt et al., 2015).

**1.2 Past studies**

There is a plethora of social-emotional rubrics nowadays as students’ social-emotional skills can aid in their positive development (Kern et al., 2016; Taylor et al., 2017). These rubrics can help students in their self-reflection of their social and emotional skills and discover which category they want to achieve due to their rigid and behaviourally-oriented criteria (Panadero & Jönsson, 2013).

Various studies have been conducted to understand SEL. Elias saw SEL as a set of competencies. Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) developed this definition as being able to recognise and manage emotions, set and achieve positive goals, appreciate the perspectives of others, establish and maintain positive relationships, make responsible decisions, and handle interpersonal situations constructively. The Collaborative for Academic, Social, and Emotional Learning (CASEL, 2005), however, categorised five cognitive, affective, and behavioural competencies that are intertwined: self-awareness (the ability to reflect on personal feelings, beliefs, and behaviours), social awareness (the ability to empathise with other perspectives, and respect foreign socio-cultural norms, and celebrate diversity), relationship skills (the ability to initiate and maintain positive connections with others), self-management (the ability to maintain self-motivation, goal setting, self-discipline etc.), and responsible decision making (the ability to make choices while taking into consideration one’s own well-being and the well-being of others).

However, there is a lack of appropriate tools to assess SEL (McKown, 2017). Improper SEL assessment can hinder educators’ goals of understanding students’ strengths and weaknesses. Subsequently, this may impact *formative assessment* - the decisions made on curriculum and instructions. Without good SEL assessment, it’s difficult to get reliable data to make informed decisions to foster students’ social and emotional developments.

Existing methods to assess SEL include direct observations, interviews, role-plays, and rating scales. However, rating scales is the most frequently used method as they are relatively efficient tools to represent summary characterisations of individuals’ observations of their own or other people’s behaviour (Crowe, Beauchamp, Catroppa, & Anderson, 2011). Rating scales function as imperfect “mirrors”, reflecting one’s social and emotional functioning (Elliot & Busse, 2004). Despite the imperfection, the reflection can still be useful with a good rating scale. Rating scales are relatively simple for the subjects to complete, and have been shown to be time-efficient and as significant as direct observations to assess social and emotional skills (Doll & Elliot, 1994).

Comprehensive reviews of SEL measures in the past have been critiqued for the lack of assessments that are specifically designed for universal screening of subjects. For example, the review conducted by the Humphrey et al. team recognised 189 measures, while the Crowe et al. team identified 86 measures (Humphrey et al., 2011; Crowe et al., 2011). Yet, these reviews did not include measures for universal screening. Hence, a critical review by Jenkins et al. (2014) included five common measures, while another review by Elliott et al. (2018), identified two more common measures. According to Jenkins and Elliott, these SEL measures can be used effectively to screen large numbers of subjects. The seven published SEL measures are:

1. Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2017)
2. Behavior Intervention Monitoring Assessment System (BIMAS; McDougal, Bardos & Meier, 2011)
3. Devereux Student Strengths Assessment (DESSA; LeBuffe, Shapiro, & Naglieri, 2009)
4. Social Academic and Emotional Behavior Risk Screener (SAEBRS; Kilgus & von der Embse, 2015)
5. Social Skills Improvement System Performance Screening Guide (SSIS PSG; Gresham & Elliott, 2008)
6. Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997)
7. Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1992)

The DESSA SEL screening measure is reportedly the only measure that aligns with the CASEL SEL competency model and is backed by psychometric evidence (Naglieri, LeBuffe, & Shapiro, 2011; Nickerson & Fishman, 2009). As there is a lack of measures that aligns with the CASEL SEL competency model, this research study adopts the universal screening measure that aligns with the CASEL model designed by Elliott et al.

1. **Methodology**

**2.1 Research study we are focusing on**

This research study aims to investigate the SEL behind collaborative learning using data derived from electroencephalography (EEG) wearables to detect brainwaves and analysis of video recordings of body movements and facial expressions. The problems set for this study will require the subjects to collaborate and solve together.

Sir James Britton believed that a student’s learning is an amalgamation of a community of students’ learnings. He suggests *natural learning* by grouping students together to develop their own culture, community, and learning. The interactions and cooperation between the students will generate a positive interdependence among the students’ learning goals. These interactions and cooperation act as a cornerstone to collaborative learning, allowing them to work together to maximise their own and each other’s learnings (Britton, 1970).

Our study draws inspiration from another study conducted by Elliot et al. in 2017. The study focuses on SEL as a vital role in education. CASEL has outlined a widely recognized framework consisting of five key components of SEL. Despite this, only a few assessments claim to measure these SEL components effectively. Their study focused on the initial validation of a new universal screening tool called the Social Emotional Learning Screening Assessment (SELA). The SELA was developed based on the CASEL framework and the existing SSIS Performance Screening Guide. As part of a larger project, experienced Australian teachers assessed 268 children from prep to year 3, providing initial user feedback and psychometric evidence for the SELA. Their findings suggest that the teacher-administered SELA aligns well with the CASEL model and offers a reliable, sensitive, and time-efficient tool for identifying students at risk socially and academically. While the results are preliminary but promising, further research is needed to replicate these findings in US schools and explore the tool's effectiveness with larger and more diverse student populations.

Using their metrics for motivation to learn as a template, we crafted our own classification system in the form of survey questions that will be elaborated in our data collection section. Our survey questions emphasise more on our research topic on collaborative learning. The metrics can be seen below:

  
Figure 1: Metrics from Elliot et al.’s study

**2.2 Data Collection**

**2.2.1 Activities**

The subjects will be tasked to carry out some activities that encourage collaborative learning. We prepared and printed out activities that range from puzzle solving like sudoku and crossword, to philosophical discussions like the “Ship of Theseus” and “Zeno’s Paradox”. Sudoku and crossword puzzles would require the subjects to discuss their unique strategies and approaches to solving them, while philosophical discussions would require the subjects to think critically and analyse others’ views and beliefs when they agree or disagree with them.

**2.2.2 Survey and Classification Algorithm**

The activities are recorded and reviewed for the subjects to fill up a survey. With the survey questions, the subjects are to reflect on their emotions and self-management during each minute of the activity. Hence, for a 10 minute activity, each subject has to fill up the survey 10 times. The survey questions are as follows:

1. I am aware of my moods and feelings.
2. I think before I act.
3. I can calm myself down when I get frustrated or upset.
4. I don’t react on the spur of the moment .
5. I paid attention, even when there were distractions.
6. I stayed calm even when others bothered or criticised me.
7. I allowed others to speak without interruption.
8. I can keep my temper in check.

Using a five-point scale for each question, we totaled up the points and classified each data point into 3 classes. For our preliminary research, the minimum total was 18 and maximum was 40. The point ranges for classes 1, 2, and 3 are 18 to 25, 25 to 33, and 33 to 40 respectively.

**2.2.3 MUSE 2 Headset**

EEG is a physiological clue in which electrical activities of the neural cells cluster across the human cerebral cortex. EEG is used to record such activities and is reliable for emotion recognition due to its relatively objective evaluation of emotion vis-a-vis non-physiological clues, such as facial expression and gesture (Suhaimi et al., 2020).

We used a mobile headset called MUSE 2 to record the subjects’ bio-signals as they carry out the activities for 10 minutes or 600 seconds. Each headset is linked to a computer to record and transfer the data into a comma-separated values (CSV) file via Window’s Powershell. One activity was recorded using the Mind Monitor application that also records the same variables and more. Each recording is 700 seconds long to ensure ample buffer time before the start of each activity. Each CSV file contains 6 columns: Timestamps, TP9, AF8, AF7, TP10 and Right AUX. To ensure that the MUSE 2 devices are used optimally, we removed any hair strands where the electrodes would be placed as they hinder the brainwave signal connections. The electrodes are AF7 (left frontal lobe), AF8 (right frontal lobe), TP9 (left temporal lobe) and TP10 (right temporal lobe). Frontal and parietal lobes seem to store the most information about emotional states (Suhaimi et al., 2020). We followed the standardized procedure for the placements of these electrodes across the skull, which also conforms to the international 10-20 EEG system for positioning of transcranial magnetic stimulation (Herwig et al., 2003).

**3 Data Processing and Analysis**

**3.1 Processing Raw EEG Data**

The subjects were instructed to blink five times consecutively together, following a 40 beats per minute metronome, to signify the start of the activity. This is done to easily detect and synchronise the data recorded on separate CSV files based on the bio-signals as blinks are artefacts that can be detected.

Using a Python code that we have constructed, we successfully detected the 5 blinks and shortened each CSV file to 600 seconds, or less if the MUSE 2 device disconnected prematurely. The Python code is designed to remove noise and extract relevant segments for further analysis by using a bandpass filter to filter out frequencies outside the 0.5-30 Hz range, common for EEG signals, and applies this filter to four EEG channels (TP9, AF7, AF8, TP10). Peaks (potential blinks) are detected in each filtered channel, and the combined peaks across channels are identified. A function searches for sequences of blinks within a specified time interval, and if found, determines the start of the experiment 1 second after the last blink in the sequence. The relevant 600-second segment of the experiment is then extracted and saved to a CSV file. Additionally, the timings of the blinks within the first 60 seconds are compiled and saved. The code also visualizes the original and filtered EEG data, highlighting the detected blinks.. This preprocessing step is crucial for ensuring clean and meaningful EEG data for subsequent analyses.



We performed Fast Fourier Transform (FFT) on the CSV files. FFT is a mathematical technique used to transform a signal from its time domain into its frequency domain. In the context of EEG data, FFT is employed to analyze the frequency components of the brain's electrical activity, which can provide insights into various neural processes.

EEG signals are complex and contain a mixture of different frequency components, each associated with different types of brain activity. By applying FFT to EEG data, we can decompose these signals into their constituent frequencies, allowing us to identify and quantify the presence and power of specific frequency bands. These frequency bands are typically categorized as:

* Delta (0.5-4 Hz): Often associated with deep sleep.
* Theta (4-8 Hz): Linked to drowsiness, meditation, and creativity.
* Alpha (8-13 Hz): Commonly related to relaxed, wakeful states and closing of the eyes.
* Beta (13-30 Hz): Associated with active thinking, focus, and problem-solving.
* Gamma (30-50 Hz): Linked to high-level cognitive functions and information processing.

By extracting these frequency components, researchers and clinicians can analyze patterns in brain activity, diagnose neurological conditions, monitor cognitive states, and even develop brain-computer interfaces. The FFT thus serves as a crucial tool in transforming EEG data into a more interpretable and actionable form, enabling a deeper understanding of brain function and dynamics.

This Python script processes EEG data stored in a CSV file to compute and extract features using Fast Fourier Transform (FFT). A function *compute\_fft\_features* is defined to calculate FFT-based features, such as the mean, variance, power spectral density (PSD), and band powers (delta, theta, alpha, beta, gamma) for a given signal. The script segments the EEG data into 60-second segments, converts the data into NumPy arrays, and iterates over each segment to compute the FFT features for each channel (TP9, AF7, AF8, TP10).

For each segment, the script stores the computed features in a dictionary, which includes the segment number and the FFT features for each EEG channel. These dictionaries are appended to a list, which is then converted into a DataFrame. The DataFrame is organized to ensure the 'Segment' column is first, followed by the FFT features. Finally, the DataFrame is saved to a new CSV file, which contains the FFT features for each segment of the EEG data. This process facilitates subsequent analysis by transforming the raw EEG data into a structured format containing key frequency-domain features.

We combined the data processed from the raw EEG data and the classifications from the self-reporting survey. The SEL class for each survey filled corresponds to the 60-second segment of EEG data. With that, we collected a total of 318 data points from 332 after cleaning the dataset.

**3.2 Feature Selection**

We calculated the correlation matrix for the dataset that displays the correlation coefficients between every pair of features in the dataset. These coefficients, which range from -1 to 1, indicate the strength and direction of the linear relationship between variables. To visualize the correlation matrix, a heatmap was generated using Seaborn, a Python data visualization library.

After establishing the correlations between the variables, we focused on identifying which features were most strongly related to the target variable, the SEL classes. To achieve this, the absolute values of the correlations with the target variable were extracted from the correlation matrix. The correlation values are as follows:

Segment/Minute 0.138425

TP9\_mean\_fft 0.006722

TP9\_var\_fft 0.231172

TP9\_psd 0.227247

TP9\_delta\_power 0.106978

TP9\_theta\_power 0.071773

TP9\_alpha\_power 0.012810

TP9\_beta\_power 0.034156

TP9\_gamma\_power 0.125425

AF7\_mean\_fft 0.139302

AF7\_var\_fft 0.147612

AF7\_psd 0.148703

AF7\_delta\_power 0.076344

AF7\_theta\_power 0.115976

AF7\_alpha\_power 0.145811

AF7\_beta\_power 0.163136

AF7\_gamma\_power 0.002383

AF8\_mean\_fft 0.056462

AF8\_var\_fft 0.262657

AF8\_psd 0.260068

AF8\_delta\_power 0.123768

AF8\_theta\_power 0.118261

AF8\_alpha\_power 0.030771

AF8\_beta\_power 0.109773

AF8\_gamma\_power 0.114700

TP10\_mean\_fft 0.168851

TP10\_var\_fft 0.300065

TP10\_psd 0.304678

TP10\_delta\_power 0.144365

TP10\_theta\_power 0.110584

TP10\_alpha\_power 0.124033

TP10\_beta\_power 0.145450

TP10\_gamma\_power 0.047698

Headset 0.065986

Classification 1.000000

However, all of the variables show little correlation to the SEL classification. Hence, we could not filter out features that did not show a meaningful correlation with the target variable. This does not mean, however, that the individual variables are insignificant and that the predictive performance is poor, it means that no single feature is a strong predictor of the SEL class. This suggests that a combination of these features is necessary to predict the class effectively.

To circumvent this issue, we performed Recursive Feature Elimination (RFE) to select the most significant features. A RandomForestClassifier was chosen as the underlying model for the RFE process. The RFE process was configured to select the top 10 most significant features (n\_features\_to\_select=10) from the dataset. The importance of each feature is determined by the classifier, and those with the lowest importance scores are eliminated. The selected features were identified and used to form the final dataset with the corresponding SEL classes to train and test our selected machine learning algorithms:

['TP9\_theta\_power', 'TP9\_gamma\_power', 'AF7\_delta\_power', 'AF7\_gamma\_power', 'AF8\_var\_fft', 'AF8\_psd', 'AF8\_theta\_power', 'AF8\_gamma\_power', 'TP10\_var\_fft', 'Headset']

**3.3 Training and Testing the Machine Learning Algorithms**

Previous research employed various methods of classifiers to classify the different types of emotional states, such as K-nearest neighbour (KNN), regression tree, Bayesian networks, support vector machines (SVM), and artificial neural network (ANN). We decided to test our dataset against the two most common machine learning algorithms for emotion-classifying research - KNN and SVM. We split the dataset 80/20 - 80% for training and 20% for testing and validation.

**3.3.1 K-nearest neighbour (KNN)**

The features are standardized using a StandardScaler to ensure that they have a mean of 0 and a standard deviation of 1, which is important for the KNN algorithm as it relies on distance metrics. The model is first trained on the training data using the default parameters, and the accuracy is evaluated on the test set. The initial accuracy score before tuning was 0.65625, indicating a moderate level of performance. The confusion matrix and classification report provide further insights into the model's performance, showing that while precision and recall are reasonable for class 2, the performance for classes 1 and 3 is less satisfactory.



To improve the model, grid search with cross-validation is performed. This method systematically explores a specified range of hyperparameters (n\_neighbors, weights, and p) to find the combination that yields the best accuracy. The grid search identified the optimal parameters as {'n\_neighbors': 17, 'p': 1, 'weights': 'distance'}. With these tuned parameters, the model's accuracy improved slightly to 0.6875.



The final classification report reveals that while the overall accuracy has improved, there are still significant differences in performance across the classes. The weighted F1 score of 0.636 suggests that while the model performs reasonably well overall, the imbalance and variability in class performance could still be an issue.

**3.3.2 Support vector machine (SVM)**

We trained an SVM model with a linear kernel using the training dataset. After fitting the model, predictions are made on the test dataset. The model achieves an accuracy of 0.6094, a precision of 0.6675, and a recall of 0.6094. These metrics suggest that the model performs moderately well but may benefit from further optimization, particularly in terms of balancing precision and recall across classes.

To improve the model's performance, a grid search with cross-validation is conducted to identify the optimal hyperparameters for the SVM. The search explores different values for C (the regularization parameter), gamma (the kernel coefficient for non-linear SVMs), and the kernel type (linear and rbf). The best parameters identified are C: 0.1, gamma: 0.1, and kernel: linear. With these parameters, the model is re-evaluated, resulting in a slight improvement in performance metrics: an accuracy of 0.6406, precision of 0.6799, and recall of 0.6406. The linear kernel, combined with the specified C and gamma, provides a better balance between bias and variance, enhancing the model's overall effectiveness.

In the final stage, feature scaling is applied to the dataset using StandardScaler to standardize the input features. This step is crucial for SVMs as they are sensitive to the scale of the input data. After scaling, the SVM model is retrained and evaluated. The accuracy remains relatively stable at 0.625, while precision improves to 0.7177, indicating that scaling has positively impacted the model's ability to classify with higher confidence. However, the recall remains at 0.625, suggesting that while the model's precision has improved, the ability to correctly identify all relevant instances remains a challenge.

The SVM model's performance is moderately improved through hyperparameter tuning and feature scaling. The grid search optimization enhanced the model's accuracy, precision, and recall slightly, while feature scaling provided a more significant boost to precision. However, the improvements were incremental, and the overall model accuracy suggests that further optimization or alternative modeling approaches might be necessary to achieve substantial performance gains.

**4 Areas for Improvement**

For future research, we can consider using a more objective metric to classify the subjects’ SEL categories. Using a self-reporting survey to classify introduces a bias to the data as the subjects may not know how to accurately determine their classifications. The five-point scale used for the survey questions are subjective and the subjects may not know where they lie on the scale. Furthermore, the subjects' reflection and recollection of their emotions may differ from other subjects as there is no standardized, objective metric for all the subjects to follow.

Subjects may also display dishonesty when filling up the survey due to feelings of embarrassment or guilt. As the activities planned focus on collaborative learning and encourages discussion, disagreements are anticipated and subjects may react negatively to these situations. Thus, when reviewing the recordings of the activities and filling up the survey questions, subjects may choose to answer dishonestly to mask their negative emotions.

We can consider the possibility of using a more objective metric like facial recognition softwares to identify emotions and classify their SEL categories, or using an objective third-party to review the video and classify the subjects’ SEL based on a standardized metric. This will minimise the biasedness, and lack of blinding, of a self-reporting survey.

Syncing one MUSE headset device to one laptop is logistically inefficient. Additionally, this introduces a synchronisation issue which we solved using the 5-blink method to signify the start of the activity, and subsequently, constructed a code to detect the 5-blink sequence. Future researchers can consider investing in an multi-input interface to connect and record multiple EEG streams at once. This can reduce the number of connections needed, and will solve the synchronisation issue.

**5 Conclusion**

In conclusion, this study explored the application of machine learning algorithms, specifically K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), for classifying social-emotional learning (SEL) based on EEG data. The findings demonstrate the potential of EEG signals as a valuable input for SEL classification, with both KNN and SVM showing moderate accuracy and the ability to capture key patterns in the data. While KNN provided a straightforward approach with relatively decent performance, SVM, especially after parameter tuning and scaling, achieved improved precision and recall, indicating its robustness in handling complex data structures. The results underscore the importance of feature selection, data preprocessing, and algorithm tuning in optimizing model performance. Future work should focus on refining these models with larger, more diverse datasets and exploring the integration of additional physiological signals to enhance classification accuracy and applicability in real-world educational settings.

**References:**

Astroth, K. A. (2017). Positive Youth Development in Diverse and Global Contexts. Journal of Youth Development, 12(3), 63-66.

Britton, J. (1970). Language and learning (p. 7). London: Penguin Books.

Crowe, L. M., Beauchamp, M. H., Catroppa, C., & Anderson, V. (2011). Social function assessment tools for children and adolescents: A systematic review from 1988 to 2010. Clinical Psychology Review, 31, 767–785.

De Fruyt, F., Wille, B., & John, O. P. (2015). Employability in the 21st century: Complex (interactive) problem solving and other essential skills. Industrial and Organizational Psychology, 8(2), 276-281.

Doll, E., & Elliott, S. N. (1994). Consistency of observations of preschoolers' social behavior. Journal of Early Intervention, 18(2), 227–238

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students’ social and emotional learning: A meta‐analysis of school‐based universal interventions. Child development, 82(1), 405-432.

Edition, C. Behavioral and Emotional Screening System (BESS). Bloomington, MN: Pearson.

Elliott, S. N., & Busse, R. T. (2004). Assessment and evaluation of students' behavior and intervention outcomes: The utility of rating scale methods. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Vol. Eds.), Handbook of research in behavioral disorders. Chapter 7. Handbook of research in behavioral disorders (pp. 123–142). New York: Guilford Publications

Frey, N., Fisher, D., & Smith, D. (2019). All learning is social and emotional: Helping students develop essential skills for the classroom and beyond. Ascd.

G. L. Ahern and G. E. Schwartz, “Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis,” Neuropsychologia, vol. 23, no. 6, pp. 745–755, 1985.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. Journal of child psychology and psychiatry, 38(5), 581-586.

Gresham, F., & Elliott, S. N. (2008). Social skills improvement system (SSIS) rating scales. Bloomington, MN: Pearson Assessments.

H. Gunes and M. Piccardi, “Bi-modal emotion recognition from expressive face and body gestures,” Journal of Network and Computer Applications, vol. 30, no. 4, pp. 1334–1345, 2007.

Herwig, U., Satrapi, P., & Schönfeldt-Lecuona, C. (2003). Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. Brain topography, 16, 95-99

Kern, M. L., Benson, L., Steinberg, E. A., & Steinberg, L. (2016). The EPOCH measure of adolescent well-being. Psychological assessment, 28(5), 586.

LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2009). Devereux student strengths assessment. Lewisville, NC: Kaplan.

Marbach, J. (2017). A Test Review: Mathematics Fluency and Calculation Tests (MFaCTs) review, by CR Reynolds, JV Voress, & RW Kamphaus.

McDougal, J. L., Bardos, A. N., & Meier, S. T. (2011). Introduction to BIMAS: Behavior Intervention Monitoring Assessment System.

McKown, C. (2017). Social-emotional assessment, performance, and standards. The Future of Children, 157-178.

Nickerson, A. B., & Fishman, C. (2009). Convergent and divergent validity of the Devereux Student Strengths Assessment. School Psychology Quarterly, 24(1), 48.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. Educational research review, 9, 129-144.

Suhaimi, N. S., Mountstephens, J., & Teo, J. (2020). EEG‐based emotion recognition: a state‐of‐the‐art review of current trends and opportunities. Computational intelligence and neuroscience, 2020(1), 8875426.

von der Embse, N. P., Kilgus, S. P., Solomon, H. J., Bowler, M., & Curtiss, C. (2015). Initial development and factor structure of the Educator Test Stress Inventory. Journal of Psychoeducational Assessment, 33(3), 223-237.

Walker, H. M., & Severson, H. H. (1992). Systematic screening for behavior disorders (SSBD). Sopris West, 1140 Boston Ave., Longmont, CO 80501.

Waters, E., & Sroufe, L. A. (1983). Social competence as a developmental construct. Developmental review, 3(1), 79-97.