Chapter 10
Measuring Multidimensional Facets of SRL Engagement with Multimodal Data

Megan Wiedbusch, Daryn Dever, Shan Li, Mary Jean Amon, Susanne Lajoie, and Roger Azevedo

Abstract Essential to achieving adaptive intelligent AI-based education systems is theoretically grounded data measurement and analysis, and the subsequent data-supported individualized interventions that foster learner-system engagement. However, engagement is a challenging psychological construct to define and measure given the variation of theoretical conceptualizations of engagement and the various facets of engagements (e.g., behavioral, emotional, agentic, (meta)cognitive, and self-regulated learning). In this chapter we (1) define and situate a multi-faceted conceptualization of engagement (based on the interrelated aspects of student engagement) within SRL, (2) introduce the integrative model of multidimensional self-regulated learning engagement to include cognitive, emotional, and behavioral facets of engagement; (3) briefly review the current conceptual, theoretical, and methodological approaches to measuring engagement and showcase how the use of multimodal data for this work has contributed to our understanding of learning in learning systems. Engagement-relevant data discussed within this chapter includes self-reports, log or behavioral streams, oculometrics, physiological sensors (e.g., skin conductance, heart-rate, etc.), facial expressions, body gestures, and think- and emote-alouds. We can leverage these multimodal data to reflect the dynamic and nonlinear nature of engagement that are frequently obfuscated by traditional unimodal methods (e.g., self-reports). However, it is crucial that when multimodal data is converged for this purpose, we consider a unifying theoretical grounding of engagement that is general enough to be applied across intelligent

M. Wiedbusch (✉) · D. Dever · M. J. Amon · R. Azevedo
University of Central Florida, Orlando, FL, USA
e-mail: megan.wiedbusch@ucf.edu; daryn.dever@ucf.edu; mary.jean.amon@ucf.edu; roger.azevedo@ucf.edu

S. Li
Lehigh University, Bethlehem, PA, USA
e-mail: shla22@lehigh.edu

S. Lajoie
McGill University, Montreal, QC, Canada
e-mail: susanne.lajoie@mcgill.ca

systems and the contexts in which they are used but specific enough to be useful in the design and development of analytical methods.; and (4) provide a methodological overview with contextualized examples to inform the research study design of future testing and validation of our integrative model of multidimensional self-regulated learning engagement using multimodal data. Our methodological overview identifies how different modalities of measurement and their temporal granularity contribute to the measurement of engagement as it fluctuates within the different phases of self-regulated learning. We conclude our chapter with an exploration of the implications of this guide as well as future directions for researchers, instructional designers, and software engineers capturing and analyzing engagement in digital environments using multimodal data.

Keywords Engagement · Multimodal data · AI-based education (AIED) systems

1 Introduction

Engagement is not just how involved a learner is with their task, but rather is goal-directed action that serves to help an individual progress academically, satisfy motivations, and create motivationally supportive learning environments (Reeve et al., 2019). Engagement has the potential to tackle persistent educational issues of low achievement (Boekaerts, 2016; Sinatra et al., 2015), risk-behaviors (Reschly & Christenson, 2012; Wang & Fredricks, 2014), and high rate of student boredom and alienation (Chapman et al., 2017; Fredricks et al., 2016, 2019a, b). However, as many researchers draw attention to, there is a notable inconsistency in both the conceptual definition and measurement of engagement (Azevedo, 2015; Fredricks et al., 2019a; Li & Lajoie, 2021). Sinatra et al. (2015) suggest that this lack of clarity derives from the construct of engagement being developed out of the assessment approach instead of grounding engagement research in a theoretical framework explicitly.

In this chapter, we address this issue by extending the Integrative Model of Self-Regulated Learning (SRL) Engagement (Li & Lajoie, 2021) to a multimodal approach for measuring the multidimensional facets of engagement. We begin by broadly defining engagement as a multifaceted construct and its relationship to self-regulated learning (SRL). Next, we introduce the extension of Li and Lajoie's (2021) Integrative Model of SRL Engagement to include additional dimensions of engagement (i.e., cognitive, behavioral, and emotional engagement). This is followed by a brief description of new underlying assumptions, strengths, and challenges of this model. We then review various unimodal methods of measuring engagement as a basis for data channel convergence and multimodal assessment. Additionally, we provide a conceptual approach of our own by providing examples of how various data channels can be used to interpret student engagement using the newly proposed model. We conclude with a discussion of limitations, future

directions, and implications for designing and developing AIEd systems that utilize theoretically grounded multimodal measures of engagement.

2 What Is Engagement?

Engagement during learning is a multidimensional construct with four distinct but intercorrelated aspects—behavioral, emotional, cognitive, and agentic—that refers to the extent of a student's active involvement in their learning (Fredricks et al., 2004; Reeve, 2012 extension of Connell & Wellborn, 1991). It is a construct that is inherently dynamic as it ebbs and flows during learning, whether that be for a single task (at the granularity of minutes) or across entire courses (at the granularity of months). When measuring and assessing a learner's quality and quantity of engagement, one must consider the level of attention and effort (behavioral engagement), the depth and quality of the strategy use sophistication (cognitive engagement), presence of facilitating and inhibiting emotions of interest and curiosity (emotional engagement), and the agency with which the learner is able to manipulate and adapt their own learning (agentic engagement). Below we briefly define these four facets.

Behavioral engagement refers to the learner's effortful involvement in their learning through strategy use and activities to stay on task via attention, effort, and persistence (Skinner et al., 2009; Reeve et al., 2019).

Cognitive engagement refers to "the extent to which individuals think strategically along a continuum across the learning or problem-solving process in a specific task" (Li & Lajoie, 2021, p. 2).

Emotional engagement refers to the presence of task-related emotions that may support or inhibit other types of engagement such as interest, curiosity, and anxiety (Reeve, 2013).

Agentic engagement refers to a learner's constructive contribution to their learning such as offering suggestions, asking questions, recommending objectives, and seeking opportunities to steer their learning (Reeve, 2013). In this chapter we do not directly address agentic engagement, however when considering the development of adaptive and intelligent systems, agentic engagement may play a vital role, especially when considering the independence of a learner who is self-regulating.

3 Extension of the Integrative Model of Self-Regulated Learning (SRL) Engagement

Just as engagement is a multidimensional construct, self-regulated learning (SRL) is also a multidimensional construct that refers to the active modulation and regulation of one's learning (see Panadero, 2017 for a review of SRL models). Researchers
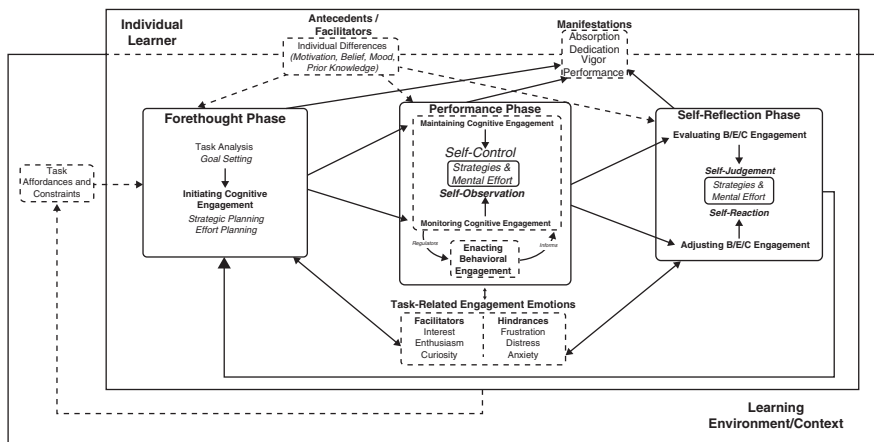
Fig. 10.1 Integrative Model of Multidimensional SRL Engagement (ISSME)

have suggested that due to the large overlap between the two constructs, we should consider integrating them (Wolters & Taylor, 2012), a call that recently Li and Lajoie (2021) responded to with their introduction of the Integrative Model of Self-Regulated Learning Engagement. This new model situates cognitive engagement inside of SRL to improve how we understand how, why, what, and when learners are more efficient and effective learners. Specifically, their model suggests that cognitive engagement fluctuates in both quality and quantity continuously throughout three sequential phases of SRL as described in Zimmerman (2000) – forethought phase (task analysis, goal setting, and strategic planning), performance phase (self-control of task strategy and self-observation), and self-reflection phase (self-evaluation, causal attribution, and adaptive self-reaction). We propose the Integrative Model of Multidimensional SRL Engagement (IMMSE) an extension of this model to include the additional facets of engagement (i.e., behavioral and emotional engagement; see Fig. 10.1).

As an expansion of the original model and its assumptions, the IMMSE places the learner (darker outline) within the learning environment or context (outermost box). Note that part of the learner is situated outside of the learning context (top dotted line). This highlights that the learner brings certain individual differences (e.g., personality, working memory capacity, prior knowledge) from the previous learning experiences such as motivations, beliefs, and moods, which will impact their current and future learning. Individual differences impact the learner both inside and outside of the specific learning context. Additionally, learners take manifestations (e.g., new beliefs, feelings of efficacy, vigor) of their learning out of the current context that will continue to impact their future learning experiences. The IMMSE suggests that those facilitators and manifestations cycle through each phase of SRL but not necessarily in a linear fashion.

In the forethought phase, a learner analyzes their task to set goals. This analysis will then initiate cognitive engagement as a learner plans how best to achieve those goals and with how much effort they should use. We show that task affordances and constraints from the environmental context such as available tools, scaffolding techniques, or pedagogical support must be considered within the forethought phase for goal setting and strategic planning. These are fed in from the learning environment or context into the individual learner. When we consider what engagement-sensitive AIEd systems will look like, we then have to consider the direct effect the learner's behavior, emotions, and cognition will have on the environment. This is depicted as the dotted line running from the leaner back into the task affordances and constraints. We discuss this feedback loop in greater detail in Sect. VI. Limitations and Future Directions.

Next, learners move into the performance phase where they cognitively regulate and monitor their behavioral engagement with the learning environment. The IMMSE argues that cognitive efficiency is a core feature of SRL engagement, such that learners' cognitive monitoring helps inform how best one should strategically and efficiently manage their cognitive engagement (Li & Lajoie, 2021). During the performance phase, cognitive engagement regulates the enactment of behavioral engagement. Behavioral engagement, according to Reeve, Cheon, & Jang, (2019), must be observable and therefore is enacted only within the performance phase. This enactment then informs cognitive engagement maintenance and monitoring mechanisms. As many of the studies that measure engagement demonstrate, cognitive engagement is usually measured by learner use of strategies. The IMMSE further extends from the original integrative model of SRL engagement (Li & Lajoie, 2021) by creating a distinction between engagement and strategy use, a commonly used proxy of cognitive engagement. Our model posits that cognitive engagement helps regulate behavioral engagement, which can help explain when behaviors indicate one level of engagement, but effort and attention indicate another. For example, behaviorally, one may appear to be reading – their eyes scanning over a page. However, when asked about what was just read, the learner may not be able to recall or even mention having their thoughts trail off. In this way, we can show there is behavioral engagement in the use of a strategy, but that there is low cognitive engagement and, in turn, low-quality engagement.

During the self-reflection phase, learners evaluate and adjust their behavioral, emotional, and cognitive engagement based on their reaction and evaluation of those strategies and the effort exerted. This is a slight expansion of the original integrative model such that, not only are learners reflecting on their cognitive engagement, but also on their emotional and behavioral engagement. For example, one might reflect that their interest (emotion engagement) is waning in a particular topic but that they recognize that topic as vital to achieving a particular learning subgoal (cognitive engagement). They determine that their approach at notetaking (behavioral engagement) has caused their declining interest. In the subsequent forethought phase after this reflection, they then plan to update their approach based on this next context and understanding of both themselves, their tasks, and current progress toward their goals.

Another expansion includes the addition of task-related engagement emotions that both facilitate and inhibit the other types of engagement throughout all of the SRL phases. Emotional engagement is unique comparatively to the other two types of engagement as it can inhibit or facilitate the amount of engagement activity that is available to the learner (Reeve et al., 2019). For example, when emotional engagement facilitators, such as an interest, are high, a learner may exert more cognitive effort than normally applied. In the IMMSE, task-related engagement emotions have a bi-directional relationship between each phase. This assumes that these emotions act as both catalysts for the internal cognitive and behavioral processes of SRL and engagement as well as products of those same processes. The IMMSE does not make any explicit assumptions about how those emotions are regulated and modified within each phase (see Harley et al., 2019).

Overall, the IMMSE suggests that engagement is an ever-changing process that fluctuates within learning. However, it is important to note that even if engagement is low quantitatively, this is not the same as disengagement. Many researchers have begun to theorize that engagement and disengagement are two distinct processes that lead to different learning consequences (Cheon et al., 2018; Jang et al., 2016; Haerens et al., 2015; Reeve et al., 2019). As such, this model does not make any distinct assumptions about disengagement but is rather focused on the temporal fluidity of engagement. This fluidity has many interconnected components that are often only ever given a cursory glance in many models of SRL such as emotions and motivations.

The IMMSE provides a theoretical grounding for future research to examine how best to measure engagement, and subsequently use those measurements for adaptive design. Additionally, it provides the groundwork for which researchers specify what it is exactly they are measuring to help clarify some of the conceptual confusion across studies. In the next sections, we review how previous work has measured engagement before providing our own approach that is grounded within this model.

4 Unimodal Methods for Studying Engagement

In this next section, we review how previous research has captured and analyzed engagement while highlighting the strengths and limitations of each approach. We follow by providing some of the new attempts at converging data channels for studying engagement. It is important to note that this review is not exhaustive in nature as some methods such as gesture recognition (e.g., Ashwin & Guddeti, 2019), teacher ratings (e.g., Fredricks & McColskey, 2012), or administrative (or institutional) data (e.g., Mandernach, 2015) have been previously used as measures of engagement but are not discussed for the sake of brevity. This section should serve as a general overview for some of the methods that have been used to measure engagement, but we direct readers to additional conceptual and systematic reviews for additional studies (Azevedo, 2015; Dewan et al., 2019; Fredricks et al., 2016, 2019a, b; Henrie et al., 2015; Li, 2021).

4.1 Clickstream Data/Log Files

Log files are sequential events or data streams where the concurrent interactions of an individual with a system (i.e., human-machine interaction) are captured (Oshima & Hoppe, 2021). Specifically, log-file data typically records the initiator (e.g., student, pedagogical agent) of an action on objects or elements within a system and what time point this action was initiated or completed. Log-file data have been used throughout literature to capture, measure, and analyze student engagement within virtual learning environments throughout a variety of domains and contexts including science (Gobert et al., 2015; Li et al., 2020), education (Henrie et al., 2015), computer science (Shukor et al., 2014), etc.

The use of log-file data can assist researchers in revealing evidence of disengaged behaviors, learner profiles of engagement, and how this relates to learners' overall learning outcomes. For example, a study by Gobert et al. (2015) utilized log files to calculate learners' frequency of actions, the amount of time between actions, and duration of the actions as they learned about ecology with a microworld. Results from this study found that using log files to measure engagement can indicate gaming the system (Baker et al., 2013) behaviors that are associated with poor learning outcomes. Similarly, log-file data recorded from an online learning platform was used to examine learner participation by utilizing the characteristics of each post to predict learners' level of engagement (Shukor et al., 2014). Results from this study found that the most effective predictors of engagement were metrics about posts where learners shared information or posted high-level messages (i.e., elaborative text).

Other studies have used log-file data to identify learner profiles of cognitive engagement during learning. Kew and Tasir (2021) analyzed log files to identify behaviors of low and high cognitive engagement displayed by learners. This study defined engagement by the quality of posts on an educational forum where each learner's cognitive engagement level was determined by comparing the ratio of low-level cognitive contributions (e.g., providing an answer to a post without explanation) to the ratio of high-level cognitive contributions to the e-learning forum (e.g., providing an explanation). Learners were identified as having either high, high-low, or low cognitive engagement depending on the relationship between the proportion of high to low cognitive engagement displayed in their posts. Findings from log-file data found that most learners were categorized as having low cognitive engagement on online forums, providing insight as to how to encourage cognitive engagement through e-learning platforms. Similarly, Li et al. (2020) identified profiles of learners based on their log-file data as they learned with BioWorld, a simulation-based training environment. Findings from latent profile analysis revealed several different types of cognitive engagement including recipience, resource management, and task-focusing (Corno & Mandinach, 1983). Additionally, results found that learners who were categorized as either resource management or task-focus cognitive engagement had greater diagnostic efficacy than learners with recipience cognitive engagement.

Log files, as demonstrated in the studies above, are revealed as important indicators and measures of engagement. Using log files has several advantages for researchers as log files: (1) are unobtrusive process-based data that can be collected online within traditional (e.g., classroom) or nontraditional (e.g., virtual) learning environments; (2) gather rich temporal data that can be contextualized, allowing for sophisticated analytical techniques to be used for examining individual learners' time-series data; and (3) should the task align with engagement theories, log files can serve as accurate identifiers of engagement during learning. However, there still exist limitations in using solely log-file data to capture, measure, analyze, and interpret cognitive engagement. Log files require interaction or physically expressed behaviors to capture engagement and as such is better suited to capture behavioral engagement rather than cognitive or emotional engagement. Historically, log files have been used to make inferences about cognitive engagement, but these inferences must be theoretically justified (Azevedo, 2015). Additionally, it is currently unknown in the literature at what time log files may be used to best capture engagement or at what sampling rate engagement indicators are unreliable or unable to be aligned with other data channels.

4.2 Eye Tracking and Gaze Patterns

Eye-tracking data refers to the experimental method of recording learners' gaze behaviors, including fixation points, saccades, regressions, and dwell times, as they engage in a task (Carter & Luke, 2020). Using eye-tracking data, researchers can identify where a learner looks, for how long a learner gazes at an area of interest (AOI; i.e., a region of an object to contextualize where a learner is looking), how often they move from one AOI to another, and the sequences a learner gazes at a battery of AOIs. Eye tracking has been used across multiple studies to capture learner engagement during reading tasks (Miller, 2015), learning with virtual environments (Bixler & D'Mello, 2016; Wang et al., 2020), designing cueing animations (Boucheix et al., 2013), etc.

Miller (2015) equated eye-tracking data, more specifically dwell times (i.e., aggregation of fixation durations) to increased thinking and attention on specified AOIs. For example, a learner who has a greater dwell time on one object would be assumed to have been thinking about the object more than an object where dwell time was lower. However, there is a large assumption being made – a learner is not engaged with material if they are not looking at the material and they are engaged if they are looking at the material. As such, studies have attempted to examine mind-wandering patterns in relation to learning outcomes using eye-tracking data. Mind wandering, also known as zoning out, is an unintentional attentional shift toward non-task-related thoughts (Killingsworth & Gilbert, 2010). Bixler and D'Mello (2016) used eye tracking to detect when a learner demonstrated mind-wandering behaviors during a reading task. During this reading task, participants were asked both during a passage and at the end of a page to report occurrences of mind

wandering while calibrated to an eye tracker. Using machine learning techniques, mind wandering was detected with 72% accuracy. This study highlights the importance of contextualizing psychophysiological data to ensure appropriate interpretations of the data are being made.

Engagement with instructional materials was similarly detected using eye tracking by D'Mello et al. (2012) and integrated with an intelligent tutoring system, Gaze Tutor, to provide learners scaffolding during a task. Individual learners' gaze battens were used to identify if the learner was disengaged to then prompt the learner via dialog to reengage the learner. Findings of this study reported increased learner attention through the use of gaze-sensitive dialogues. Similarly, Bidwell and Fuchs (2011) identified individual learners as either engaged or disengaged using eye tracking where learners were classified into one of three states: engaged, attentive, or resistive. However, when compared to expert human coders' classification of student engagement, hidden Markov models were only 40% accurate in classifying learners, perhaps highlighting the role and impacts of subjectivity of subjects and observers in some research.

These studies highlight both the strengths and limitations of using eye tracking to identify and measure engagement. The method of collecting eye-tracking data can be expensive and intrusive to the learner due to the calibration and equipment setup. Additionally, the collection can be complex as it may be affected by the individuals' physical actions such as sweating or moving (Henrie et al., 2015). Although collecting this data can be difficult, eye-tracking data can be collected at multiple levels of granularity, from milliseconds to hours of aggregation and timespans. In addition, eye tracking can measure temporal sequences of actions through saccades, attention allocations via fixation durations and dwell times, and cognitive effort through pupil dilation, providing researchers with richly quantified dataset contextualized to the learning task (e.g., see Dever et al., 2020; Taub & Azevedo, 2019; Wiedbusch & Azevedo, 2020).

4.3   Audio/Video (Think and Emote-Alouds, Observations, and Interviews)

Audio and video serve as methods for collecting think-alouds (i.e., concurrent verbalizations) of learners' thoughts as they complete a task (Ericsson & Simon, 1984), emote-alouds where verbalizations consist of emotions, observations of learners' actions during a task, and interview data for post-task qualitative analysis (D'Mello et al., 2006). Data from audio and video can provide insight as to how learners demonstrate engagement with material during a task and provide critical contextual cues needed to make accurate inferences about engagement. For example, Tausczik and Pennebaker (2010) argue that word count calculated through think-aloud audio data can identify a learner who is dominating a conversation with a peer, teacher, or tutor as well as the level of engagement that is demonstrated by a learner (e.g., high, low).

Past studies have examined think-alouds across contexts and domains to identify engagement. For example, one study used a combination of interviews and video as learners completed math lessons to measure cognitive engagement (Helme & Clarke, 2001). Findings revealed that cognitive engagement was accurately identified through both linguistic and behavioral data from audio and video data respectively. Linguistic indicators of cognitive engagement included verbalization of thinking, seeking information, justifying an argument, etc., where behavioral indicators were primarily identified through gestures. Another study used audio data to identify linguistic matching during a negotiation between multiple parties (Ireland & Henderson, 2014). Within this study, lower task engagement levels were associated with an increase in language use and style matching (i.e., percentages of words in various linguistic categories) but were indicators of higher social engagement. More specifically, the mimicry of verbal and non-verbal communication showed an increased attention to social cues but had a negative relationship with task engagement as pairs were more likely to hit conflict spirals and impasses. A study by Ramachandran et al. (2018) also examined social engagement via audio data where the word count and the number of prompts were recorded as a conversation took place between a learner and a robot tutor. Specifically in this study, learners were required to think aloud and while doing so, a robot tutor would prompt the learner to consistently think aloud Using a robot-mediated think-aloud showed improvements in students' engagement and compliance with the think-aloud protocol compared to using just the robot without think-aloud prompting, tablet-driven think-aloud prompting, or neither the robot or think-aloud prompting, indicating the potential value of using social robots in education for (meta)cognitive engagement.

While audio and video data can serve as a non-intrusive method of rich data collection to measure cognitive, behavioral, task, and social engagement, there exist several limitations that are specific across different data collection methods (Azevedo et al., 2017). Observational methods can be expensive and require trained and paid professionals. For example, the BROMP coding technique (Baker et al., in press) is a momentary time sampling method in which trained certified observers record student's behavior and affect in a pre-determined order using an app that can then automatically apply various coding schemes. In addition, think- and emote-alouds require learners to be able to accurately and consistently verbalize their thoughts, emotions, and cognitive processes which may slow performance as they try to complete a task sometimes complex in nature (e.g., problem solving, learning about a difficult concept). Further, despite the density of utterances, these studies using these methods tend to have smaller sample sizes, making it difficult to generalize to other studies. However, audio and video data can focus on the activity level, provide qualitative aspects of engagement (e.g., emotional engagement), and contextualize other data type measures of engagement such as eye-tracking data. They also focus on the veracity of the data such that while the number of subjects may be more limited, the depth of the data collected is rich and offers a valuable corpus that can then be inspected from multiple vantages.

4.4   *Electrodermal Activity and Heart Rate Variability*

Electrodermal activity (EDA) data manifests from changes in learners' topical electrical conductance, quantifying sweat gland activity to identify stimuli such as cognitive engagement (Posada-Quintero & Chon, 2019; Terriault et al., 2021). Heart rate variability (HRV) measures the fluctuation of the duration between heartbeats to identify the temporal relationship and changes in sympathetic and parasympathetic effects on heart rate (Appelhans & Luecken, 2006). Both physiological data channels aim to mitigate the limitations of traditional techniques such as survey-based measures that can be time-consuming and cognitively demanding for the learner (Gao et al., 2020). Because of this, studies have attempted to understand how non-invasive EDA and HRV data collection methods can be used to capture and measure learner engagement.

A study by Gao et al. (2020) explored how learners' cognitive, behavioral, and emotional engagement could be captured by EDA metrics and which of those metrics are the most useful in predicting learner engagement as well as differentiating between the three types of engagement. Results from this study found that cognitive, behavioral, and emotional engagement level during class instruction can be detected with 79% accuracy across 12 EDA metrics in addition to other physiological metrics (i.e., photoplethysmography, accelerometer). In examining the relationship between EDA peak frequency and the three types of engagement, Lee et al. (2019) found that a greater number of peaks indicating increased arousal was related to greater cognitive and behavioral engagement. However, in relating EDA peak frequency to emotional engagement, the study did not find significant associations possibly due to those activating emotions, either positive or negative, that can have both positive and negative relationships with emotional engagement (Lee et al., 2019). In contrast, Di Lascio et al. (2018) found that when measuring emotional engagement during class, increased levels of arousals were related with greater levels of emotional engagement.

As seen in the slight variation in findings across studies, collecting EDA and HRV data can be challenging due to the limitations presented for data collection, analysis, and implications. Specifically, both data channels can be intrusive due to the instrumentation of learners that must occur. While some instruments, like a smart watch, can unobtrusively collect this information, more sophisticated and expensive instruments allow for greater accuracy (e.g., greater sampling rate; Henrie et al., 2015). Additionally, many precautions and considerations must be taken in both the environmental conditions (e.g., temperature) and participants' individual physiological and lifestyle differences (e.g., weight, caffeine and medication consumption, etc.; Terriault et al., 2021). Interpreting arousal via EDA and HRV data can be difficult without the use of additional data channels such as within the study by Di Lascio et al. (2018) who compared arousal data against self-report measures to triangulate the validity of arousal measures and implications. While these data channels can be used to accurately predict levels of cognitive and behavioral engagement in learners during a task or lecture, emotional engagement has yet to be

concretely identified through these techniques. However, EDA and HRV methods collect rich, fine-grained data that allow researchers to create individualized models of engagement.

4.5   Self-Reports and Experience Sampling

Self-reports and experience sampling have been long-standing measures of cognitive, behavioral, and emotional engagement due to the ease of administration and the ability to understand learners' reflections on their engagement. To obtain these data, learners are asked to report experiences and understanding of their own degree of engagement during a learning task either prior to (e.g., "Before a quiz or exam, I plan out how I will study the material"; Miller et al., 1996) or after (e.g., "To what extent did you engage with the reading material?") their task. Several studies have not only developed scales for engagement (e.g., Appleton et al., 2006; Vongkulluksn et al., 2022) and examined these scales for reliability and accuracy (Fredricks & McColskey, 2012), but have also examined and determined learners' level of engagement using self-reports and experience sampling. A study by Salmela-Aro et al. (2016) used the experience sampling method of short questionnaires throughout a science class across 443 high school students. From this sample and using latent profile analysis, this study found four profiles of learners – engaged, engaged-exhausted, moderately burned out, and burned out. Through this assessment and methodology of data collection, this study was able to examine both positive and negative aspects of engagement. Xie et al. (2019) also used experience sampling to measure cognitive, behavioral, and emotional engagement across several self-report measures of engagement. Findings from this study established event-based sampling as a more accurate way that cognitive, behavioral, and emotional engagement can be captured by self-reports. Finally, experience sampling type of self-reports allowed researchers to have a deeper exploration of how engagement relates to learner behaviors.

Using self-reports and experience sampling methodologies to capture, collect, analyze, and interpret cognitive, behavioral, and emotional engagement demonstrated by learners has several strengths The method is easy and cheap to administer to learners and provides a representation of learner reflection and perception of engagement during a task (Appleton et al., 2008). Additionally, these methods can be used to compare across scales and, as Appleton et al. (2006) argue, can be the most valid measure of both cognitive and emotional engagement as both constructs rely on learners' self-perception. However, a review of cognitive engagement self-report measures by Greene (2015) showed that researchers have begun to over rely on the information provided by these measures, without regard to the several limitations these metrics pose. For example, self-report measures of engagement have not fully developed the definition and multidimensional conceptualization of cognitive, behavioral, and emotional engagement, leading to a divided field regarding the indicators of engagement during learning (Fredricks & McColskey, 2012; Li et al.,

2020). Additionally, the assumption is made due to self-reports that engagement is stable across time, can be aggregated and misaligned with real-time behaviors demonstrated by learners, and can be measured outside of the immediate learning task (Greene, 2015; Greene & Azevedo, 2010; Schunk & Greene, 2017). One-way studies have attempted to rectify this limitation is through the prompting of self-reports throughout the learning task. However, this prompting can be disruptive to the learner as well as cognitively demanding during a learning task (Penttinen et al., 2013). As such, several pieces of literature have indicated self-reports (generally) as poor indicators of the construct that was intended to be measured (Perry, 2002; Perry & Winne, 2006; Schunk & Greene, 2017; Veenman & van Cleef, 2019; Winne et al., 2002).

4.6   Facial Expressions

Facial expressions have primarily been used to identify learners' internalized and temporal emotions as they complete learning tasks. To do so, video clips of learners are captured and enumerated using several different algorithms which identify different states of emotions including happiness, anger, joy, frustration, boredom, etc. One example is the Facial Action Coding System (FACS; Ekman & Friesen, 1978) which maps action units, or specific landmarks, onto the learner's face to monitor and quantify which facial structures move, when they move, and in conjunction with other action units. From this, emotion scores are derived which indicates the probability of an emotion being present.

Several studies have used machine learning techniques on learners' facial expressions to identify at what point of time and the duration a learner demonstrates engagement on a learning task (Grafsgaard et al., 2013; Taub et al., 2020). A study by Whitehill et al. (2014) examined methods to automatically detect instances of engagement using learners' facial expressions in comparison to human observers judging emotions displayed by learners in 10-second video clips. Findings of this study established machine learning as a valid technique for reliably detecting when a learner displays high or low engagement. Similarly, Li et al. (2021) employed supervised machine learning algorithms to identify how learners demonstrated cognitive engagement using facial behaviors as they deployed clinical reasoning in an intelligent tutoring system. Results found that three categories of facial behaviors (i.e., head pose, eye gaze, and facial action units) can accurately predict learners' level of cognitive engagement. Moreover, there were no significant differences in the overall level of cognitive engagement between high and low performers. However, learners in this study who were classified as high-performance demonstrated greater cognitive engagement as they completed deep learning behaviors.

Prior literature has shown that engagement can be detected and predicted using learners' facial expressions. However, using facial expressions to identify engagement assumes that all emotions are depicted by facial expressions. More specifically, we ask the question: do all emotions need to be expressed facially to exist?

From this, there is a limitation to determining the level and type of engagement as emotions could potentially be completely internal without outward indicators of their presence. Using facial expressions as a measurement of emotion assumes facial expressions are universal while ignoring potentially important social constructs (e.g., culture, positions of power, dynamics of relationships, etc.) and context. For example, a smile might not always indicate happiness if following bad news. In this case, the smile may be interpreted as an emotion-regulatory strategy or dismissive strategy to negative emotions. It may also indicate mind-wandering or disengagement if the smile is not related to any event or trigger from the environment or learning task.

Despite these limitations, facial expressions have been shown to be accurate and reliable indicators of engagement according to past studies (e.g., Grafsgaard et al., 2013; Li et al., 2021; Taub et al., 2020; Whitehill et al., 2014). Facial expressions, in addition to reliable detection, are non-invasive and able to be automatically coded in real time without the utilization of human resources.

4.7   EEG

Electroencephalogram (EEG) is a physiological measure of summed postsynaptic potentials as neurons fire that provide temporal information about dynamical changes in voltage as measured via electrodes attached to the scalp and a reference electrode (Gevins & Smith, 2008). From these electrode voltage signals, there are several frequency bands that have been used to measure cognitive states and processes such as vigilance decline (Haubert et al., 2018), information processing (Klimesch, 2012), and mental effort (Lin & Kao, 2018).

Pope et al. (1995) developed an engagement index based on a ratio between the amplitudes of beta, alpha, and theta frequency bands that was found more sensitive to changes in cognitive workload demands than other indices. This index has since been reconfirmed as a sensitive measure to cognitive engagement during various cognitive lab-based tasks (Freeman et al., 1999; Nuamah & Seong, 2018), and has been used to study engagement in children during reading (Huang et al., 2014), employees in workplaces (Hassib et al., 2017a), and university students during lectures (Hassib et al., 2017b; Kruger et al., 2014).

Studies using EEG are less intrusive than other brain-scanning methods and have been conducted within lectures using headsets (e.g., Kruger et al., 2014). These studies provide temporally rich and fine-grained data that are prime for measuring cognitive engagement but are computationally and resource intensive. Additional types of non-intrusive brain-scanning methods (e.g., fNIRS) have been used to detect features of engagement (e.g., Verdiere et al., 2018) using similar operationalizations of cognitive engagement which may prove to be a synergistic measurement tool to EEG in future work examining other dimensions of engagement.

4.8 Convergence Approaches

Above we reviewed several unimodal data channels that have been used to study engagement, with many of those studies using a second data channel to validate newer measures (e.g., Bixler & D'Mello, 2016; Gao et al., 2020; Taub et al., 2020). As important, the use of multiple data sources also (1) provides complementary information when used in conjunction with one another (Azevedo & Gašević, 2019; Azevedo et al., 2017; Sinatra et al., 2015), (2) provides contextual information for interpretation (e.g., Järvelä et al., 2008), and (3) can be used to develop more holistic models by identifying interrelations among related variables (Papamitsiou et al., 2020).

For example, Dubovi (2022) measured the emotional and cognitive engagement of learners using a VR simulation of a hospital room using facial expressions, self-reports, eye-tracking, and EDA. It is important to note, however, that in this study these channels were all analyzed independently. That is, these two metrics were not combined into a single "emotional engagement" metric but rather the authors report that facial expressions were used to examine dynamical fluctuations in fast-changing emotions while the self-report measured the intensity of emotions at set times throughout the task. This study shows one way to use multiple data channels to complement one another during interpretation.

Attempts have also been made to evaluate engagement using multimodal data that are used in conjunction with one another to help provide additional context. For example, Sharma et al. (2019) used both head position and facial expression to classify learners' engagement level. Their system begins by detecting the face and head position to determine the learners' attentional state (i.e., distracted or focused), and if the learner is focused, the dominant facial emotion is measured, and an engagement level is calculated based on the dominant emotion probability and a corresponding emotion weight. This value corresponds to a classification of "very engaged", "nominally engaged", and "not engaged". This approach underscores how one data channel can be used to provide contextual information for another data channel measurements to occur.

Finally, other studies have attempted to fuse multimodal data to develop more holistic measures of engagement. Papamitsiou et al. (2020) were able to use log files, eye-tracking, EEG, EDA, and self-report data in a fuzzy set qualitative comparative analysis approach (using 80%, 50%, and 20% thresholds for degree of membership) to create a multidimensional pattern of engagement. Their approach identified 9 configurations of factors to help explain performance and engagement on a learning activity. Their findings showcase how multimodal data fusion suggests more than one pattern of engagement that facilitates higher learning outcomes. That is, because engagement is a multidimensional construct, there are likely multiple avenues by which engagement impacts learning and that models of the measurements of engagement should reflect this.

The promising shift toward multimodal approaches for measuring engagement is largely driven by new analytical techniques such as those described above. However,

there are limitations around using a multimodal approach. We discuss these limitations in further detail later in this chapter (see Sect. VI. Limitations and Future Directions) but highlight that these studies are not easy to collect for a large number of participants. As such, many of these studies have relatively smaller sample sizes that should be considered when discussing generalizability across learning context. As new multimodal analytical techniques emerge, we can expect to see a large increase in these types of studies which will help address this concern. It is imperative that this work, however, be theoretically driven to avoid data hacking or phishing expeditions. Grounded within this model.

5 Theoretically Grounded Approach for Measuring Engagement with Multimodal Data

As we have previously shown, there have been many approaches to studying engagement. However, the data that currently has been used to measure engagement have varying conceptualizations and degrees of utilizing the multifaceted definition of engagement we outlined previously. This means that a direct comparison of these methods' effectiveness and efficacy is not only difficult but also ill-advised. Instead, we suggest future research needs to be explicit in what components and facets of engagement are of interest to help advise which channels would be deemed most appropriate. That is, there is likely not a single channel that will provide a single metric for best quantifying engagement. Rather, each channel has its strengths and limitations for each component and facet of engagement that must be considered. These considerations can include which phase of SRL is of interest (i.e., forethought, performance, or self-reflection), the facet of engagement that is being measured or inferred (e.g., cognitive versus behavioral versus emotional), the temporal granularity (e.g., changes in engagement moment to moment versus sub-goal to sub-goal versus day to day), the context as constrained by the environment, and combination of converging data channels (see Azevedo et al., 2017, 2019). Using multimodal data to measure psychological constructs is not a novel approach, as we have highlighted in several studies above. Often, multiple channels are used to validate another (e.g., using self-report data to validate EDA fluctuations of arousal indicating higher engagement). However, multimodal data can also be used in conjunction to provide complementary information for measuring engagement as well (D'Mello et al., 2017). Importantly though, how best to integrate and utilize multiple channels is not yet fully understood and requires additional empirical work that is grounded in a unifying model of engagement.

Below, we provide examples of how these methods could be used specifically to interpret not only each facet of engagement, but also their individual components as situated in our extended integrative model of SRL engagement (see Table 10.1). This table highlights how no one channel can capture all components of all facets of engagement. For example, facial expressions might provide fine-grained data on

Table 10.1 Data channels and examples for measuring engagement within each phase of SRL

Scenario: A learner is using an intelligent tutoring system to learn everything they can about Topic X given a specific time constraint (e.g., see Azevedo et al., 2022).

Data Channel	SRL phase and components	SRL facets	Examples	Exemplar citations (utilizing the data channel)
Eye-tracking	Forethought	Initiating cognitive engagement (strategic planning, effort planning)	Gaze patterns are used to determine how the learner scans the navigation toolbar which lists the various tools available to use (e.g., a notepad, quiz, video tutorials).	Antonietti et al. (2015), Boucheix et al. (2013), D'Mello et al. (2017), Duchowski (2007), Miller (2015) and Van Gog and Jarodzka (2013)
	Performance	Enacting behavioral engagement	Dwell times are used to determine how long the learner spends looking at the video tutorial.	
	Self-reflection	Evaluating B/E/C engagement	Gaze on relevant performance or learning metrics following the learning session provided by the system are used to determine if the learner is reflecting on their previous performance phase.	
Log files/click streams	Forethought	Initiating cognitive engagement (strategic planning, effort planning)	Event markers that are quick in succession are used to determine when and which tools the learner opens up to scan to determine what is available to use as they orient to the learning environment.	Azevedo et al. (2015), Bernacki et al. (2012), Gobert et al. (2015), Henrie et al. (2015), Järvelä et al. (2008), Kew and Tasir (2021), Li et al. (2020, 2021) and Shukor et al. (2014)
	Performance	Maintaining cognitive engagement	Dwell times are used to determine how long the learner spends looking at the video tutorial	
		Enacting behavioral engagement	Keylogging captures the notes a learner's take as they watch a video tutorial.	
	Self-reflection	Evaluating B/E/C engagement	Event markers that are quick in succession to previously visited pages or personal notes are used to evaluate the past performance phase.	
		Adjusting B/E/C engagement	State transitions inferred via modeling approaches (e.g., Markov chains) or event analysis may indicate points of reflection requiring intervention from the learner.	

(continued)

Table 10.1 (continued)

Scenario: A learner is using an intelligent tutoring system to learn everything they can about Topic X given a specific time constraint (e.g., see Azevedo et al., 2022).

Data Channel	SRL phase and components	SRL facets	Examples	Exemplar citations (utilizing the data channel)
Facial expressions	Manifestations	Absorption, dedication, vigor, performance	The change in frequency or rate of actions a learner takes may indicate changes in the level of effort that are being put forth. Additionally, log files can be used to see specific steps taken in assessment-based questions to quantify performance.	
	Antecedents/facilitators	Mood	Evidence scores for the level of joy or frustration are captured prior to the learning session to determine how a learner feels prior to interacting with the environment.	Grafsgaard et al. (2013), Li et al. (2021), Taub et al. (2020), Whitehill et al. (2014)
	Task-related emotions	Interest, enthusiasm, curiosity, frustration, distress, anxiety	Evidence scores for frustration and confusion are captured throughout the learning session to track how a learner feels at various stages of learning.	
Emote Alouds	Antecedents/ facilitators	Mood	The frequency of verbalizations where the learner indicates their mood ("I feel good today!") is calculated	Baker et al. (2013) and Craig et al. (2008)
	Task-related emotions	Basic academic-related emotions (frustration, confusion, anger, joy)	The frequency of verbalizations in which a learner indicates an emotion is calculated. ("I am frustrated I don't understand topic X!")	

Scenario: A learner is using an intelligent tutoring system to learn everything they can about Topic X given a specific time constraint (e.g., see Azevedo et al., 2022).

Data Channel	SRL phase and components	SRL facets	Examples	Exemplar citations (utilizing the data channel)
Think Alouds	Antecedents/ facilitators	Motivation, belief	The frequency of verbalizations indicating how motivated a learner is calculated. ("I am really excited to learn about topic X, it has always been a curiosity of mine.")	Azevedo et al. (2022), Appleton et al. (2006), Dent and Koendka (2016), Duffy and Azevedo (2015), Fredricks and McColskey (2012), Greene et al. (2015), Helme and Clarke (2001) and van der Graaf et al. (2022)
	Forethought	Task analysis (goal setting)	The frequency of verbalizations of explicit goals is calculated. ("I want to learn specifically about subject Y within topic X.")	
		Initiating cognitive engagement (strategic planning, effort planning)	The frequency of verbalizations of plans is calculated. ("first, I'll see what videos are available since I'm new at this.")	
	Performance	Maintaining cognitive engagement	The duration of pauses in between verbalizations of effort is calculated to compare over time.	
		Monitoring cognitive engagement	The frequency of verbalizations of effort is calculated. ("this practice problem is requiring me to try really hard.")	
	Self-reflection	Evaluating B/E/C engagement	The frequency of verbalizations of reflection is calculated. ("I think the note-taking tool was not very helpful for this task.")	
		Adjusting B/E/C engagement	The frequency of verbalizations of changes to levels of interest is calculated. ("I'm not as interested in topic X as I used to be.")	

(continued)

Table 10.1 (continued)

Scenario: A learner is using an intelligent tutoring system to learn everything they can about Topic X given a specific time constraint (e.g., see Azevedo et al., 2022).

Data Channel	SRL phase and components	SRL facets	Examples	Exemplar citations (utilizing the data channel)
Self-reports	Antecedents/facilitators	Motivation, belief, mood	Goal orientations are measured prior to learning.	Fredricks and McColskey (2012), Greene (2015), Renninger and Bachrach (2015), Vongkulluksn et al. (2022) and Wolters (2004)
	Self-reflection	Evaluating B/E/C engagement	Retrospective confidence judgments are used to measure a learner's behavioral engagement use.	
	Task-related emotions	Interest, enthusiasm, curiosity, frustration, distress, anxiety	Scales for curiosity are administered after a new sub-goal is indicated and used to determine changes in task-related emotions	
	Manifestations	Absorption, dedication, vigor, performance	Scores on a quiz after sub-goal completion are administered throughout the learning session to examine fluctuations over time	
Eda	Performance	Maintaining cognitive engagement	The frequency of high arousal peaks during video lectures is compared to determine which videos were most emotionally engaging.	Dubovi (2022), Gao et al. (2020), Di Lascio et al. (2018), Lee et al. (2019), McNeal et al. (2020) and Terriault et al. (2021)
	Manifestation	Absorption, dedication, vigor	Increases in percent change of skin conductance are used to infer changes in vigor.	
Heart rate	Antecedents/facilitators	Mood	Heart rate variability is used to infer emotional states that are captured prior to a learning session to serve as a baseline for a learner's current mood.	Gao et al. (2020) and Wang and Cesar (2015)
	Performance	Maintaining cognitive engagement	The fluctuations in heart rate variability are used to determine changes in arousal levels during various strategy uses.	
	Task-related emotions	Interest, enthusiasm, curiosity, frustration, distress, anxiety	Heart rate variability is used to infer emotional states via arousal during learning.	

one's expressed task-related engagement emotions as they fluctuate with their inter-actions between all SRL phases, but they would provide little to no information about behavioral engagement (albeit context may be inferred from the facial expressions to provide some interpretation of behaviors such as why someone might be engaging in a particular strategy).

This table highlights a couple of interesting challenges when working with multimodal data that go beyond what has already been published (see Azevedo et al., 2017, 2019, 2022; Järvelä & Bannert, 2021; Molenaar et al., 2022). First, many of these data are starved of qualitative information that can be derived from another channel. For example, this table describes an example of how log-file data could be used to evaluate engagement during the self-reflection phase. We suggest that event markers that are quick in succession to previously visited pages or work can be used to indicate reflection. Additionally, this table highlights how events are still being examined primarily independently of one another instead of thinking of actions or events that are more communal. That is, much in the same way we can use collections of facial landmarks to detect faces and facial expressions, log-file event markers could be used to create constellations indicating various types of engagement. However, unaccompanied by think-alouds or self-reports, what the learner was consciously reflecting on may not be differentiated from unconscious reflection of what was being reflected upon. That is, the learner could be consciously reviewing the length of their notes or determining if they had seen all of the material by flipping through the informational pages of an environment, but unconsciously evaluating how much effort taking those notes or reading all of that material took and whether or not they felt it was effective to their current judgment of their learning. Log files would not be able to make this distinction alone, but rather must be inferred (perhaps based on the type of content being reviewed or the order the content is reviewed). The addition of think-aloud data might provide more context as to the *why* of the reflection.

This table also highlights where the various channels benefit from the contextualization of other channels. That is, the same metrics might be recorded and not be able to provide interpretable delineation between the SRL phases without other data channels. For example, our table suggests that heart rate variability can be used to determine task-related emotions and the maintenance of cognitive engagement. However, it is important to note that this requires a level of inference-making as the metric is reporting on arousal. Additional context is needed to understand if fluctuations in the heart rate variability are driven by changes in effort (indicating performance-phase cognitive engagement maintenance) or changes in task-related emotions (e.g., anxiety or distress). By introducing additional data, such as speech or self-reports, important distinctions can be made. That is, the addition of self-report data might provide more context on the *when* or *what SRL phase* of physiological data.

As we consider multimodal data channels for inferring engagement levels of learners, we must also consider not only their temporal granularity in relation to one another (i.e., eye-tracking versus self-report measures) but also within each channel (Azevedo & Gašević, 2019), and potentially in the "fused" channels. For example,
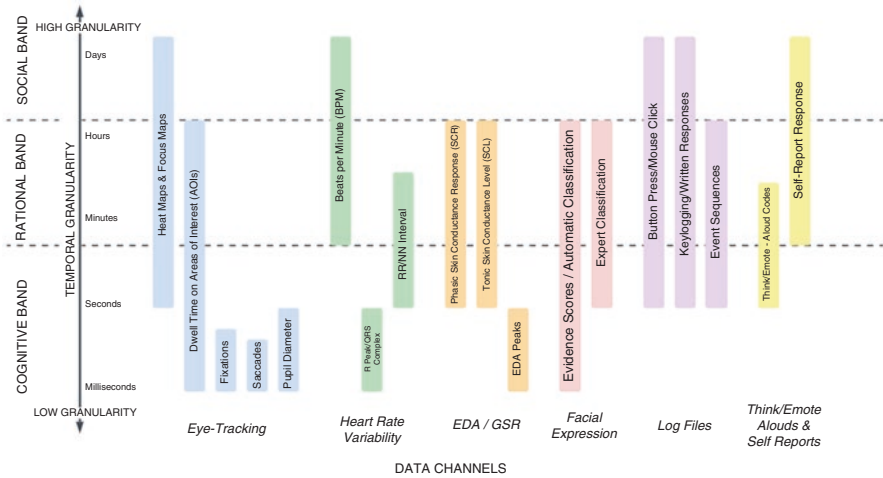
Fig. 10.2 Temporal banding of multimodal data channels for measuring engagement

within eye-tracking, we must consider the inferential implications of using more fine-grained data (e.g., fixation durations) compared to more aggregated forms of data (e.g., heatmaps). This granularity is further explored in Fig. 10.2, which is a non-exhaustive set of metric examples that can be used to make inferences about engagement. These data can be used to explore when, how long, how often, and the shape or topography of occurrences transpire during the various fluctuations of engagement within the SRL phases outlined in our model.

Within each data channel (horizontal axis), we provide examples of metrics (colored bands) that can be collected (e.g., timestamps of button presses from log files) or generated (e.g., Think/Emote-aloud codes). These have been situated along the vertical-hand's scale of more fine-grained and typically raw data up to more aggregated data. For example, fixations and saccades are collected within the millisecond range, but can be accumulated and aggregated across minutes or hours to determine dwell times. This figure outlines the relative temporal banding of the exemplar metrics that can be used to make inferences about learners' cognitive, behavioral, and emotional engagement during SRL. These inferences are best captured with multimodal data using a variety of data channels (e.g., eye-tracking) with multiple metrics (e.g., fixations, dwell time), and variables (e.g., fixation frequency, dwell time duration) that can be extracted to evaluate engagement while learning. As research continues to develop novel and innovative approaches to measuring various psychological constructs using both online trace data and offline sources, the metrics available for use are a growing list (see Darvishi et al., 2021). Those outlined above just scratch the surface at what has been previously examined, but we acknowledge that many more metrics exist that could fit well into our model. Additionally, within each of those metrics, there are many variables that can be extracted. Almost all metrics can be analyzed using frequency (i.e., how often something occurs),

duration (i.e., how long it took to occur), and their timing within the learning context/timeline (i.e., when something occurs).

The dimension of time scale of grain size has further been split (horizontal dashed lines on across the vertical axis) separating the cognitive, rational, and social bands on the left vertical) according to Newell's (1994) levels of explanation corresponding to the time scale of human actions, to highlight which data are most appropriate when making inferences about cognitive (unit tasks, operations, and deliberate acts), rational (task level activity), or social (e.g., course-length engagement) activities. This work was in large part to help the development of cognitive architectures, and as such these bands represent qualitative shifts about the type of processing assumed to occur within them and the manner researchers talk about their internal levels from a systems-level perspective (West & MacDougall, 2014). Briefly, the cognitive band represents symbolic information processing, the rational band represents the level knowledge becomes abstract to create a (imperfect) knowledge level system, and the social band refers to distributed multi-agent processing.

However, it is important to note that the delineations between the bands are not hard boundaries but rather gradual guidelines (as indicated by being dashed and not solid). Furthermore, according to Anderson's (2002) "Decomposition Thesis", there is much evidence that suggests human action occurring at grander time scales are composed of smaller actions at shorter time scales. That is, most of what occurs in the social band involves a great degree of rational and cognitive processing. This figure highlights how it is important in the work around engagement that we must be clear about the temporality of interest when discussing the quality and quantity of engagement. For example, are we concerned with fine-grained attentional shifts within a single task or the overall level of interest and emotional investment of a semester-long course consisting of multiple lectures each with multiple tasks?

6 Limitations and Future Directions

This chapter provides groundwork for future engagement research by drawing new connections between associated constructs and measures. Specifically, the narrative offers a multifaceted theoretical conceptualization of engagement and associated data channels. Considering broader implications and future directions for engagement-sensitive AIEd systems, we see an additional strength of our model (Azevedo & Wiedbusch, 2023), in that the interaction between the individual learner and the environment is one that allows for feedback loops. These have the potential to become externalized and therefore are amendable to IMMSE analysis. For example, we can imagine a system that detects waning cognitive and behavioral engagement during strategy use within the performance-phase based upon eye-tracking and log file data. Upon this detection, the system may then choose only at that moment to interrupt the learner to probe about their current emotional engagement levels and offer suggestions how best to increase levels of interest or curiosity. In this way, the system is directly adapting to the user. However, we must remember

these systems should also be used to scaffold learners, so some intrusions need not be only measurement-related in nature. These intrusions can be intervention-driven and serve as additional data-rich sources for future measurement without this being their main intent. Additionally, as the user continues to interact with the system, we can imagine that it begins to track which interventions prove to be the most successful in increasing engagement. These interventions are then made more readily available for the user within the system while also suppressing those interventions that have been shown to decrease the individual's level of engagement. In this way, the learner is directly adapting the task and environment affordances and constraints to improve their learning experience. Due to learner individual differences, these changes could be made in such a way that no two learners' environments are the same.

Future work can also test and elaborate on specific connections forwarded in the IMMSE, including how prior knowledge, task constraints, and goal setting influence engagement facets. For example, the present model (Fig. 10.1) highlights connections between task analysis (esp. goal setting) and the initiation and maintenance of cognitive engagement, which can be measured via gaze fixations and EDA, among others. Connections such as these can be empirically examined, not only to test the model, but to forward appropriate SRL interventions and measures of cognitive engagement. Additional work can also more thoroughly address the role of agentic engagement within the context of self-regulated learning, which should be expanded upon in future model iterations.

This work also recognizes the general advantages and disadvantages of multimodal approaches to assessing engagement, as well as the need for ongoing research in this broad area. A conclusion one can draw from the IMMSE model, which highlights a tension between collecting as much as possible, and knowing which channels are most helpful to a particular context and analysis, is that the model will encourage more research that contrasts the relative utility of different measurement channels and metrics – solo and combined – in studying particular constructs in particular contexts (e.g., Amon et al., 2019, 2022). For example, it is increasingly popular in the realm of multimodal measurement as a sensory-suite approach to research, where all available measurement channels are utilized during research studies within a given lab, even if a particular measurement channel is not central to the motivating research questions. Research is conducted in this fashion for good reason: Research is expensive and time-consuming, and – for those fortunate enough to afford such setups – elaborate sensory suites provide more "bang for the buck." By capturing as much information during a study as possible, researchers can push creative research questions to the forefront and harvest data for years to come. Certainly, the sensory-suite approach is a good investment in many cases, but it has some caveats. Researchers may put the cart before the horse in terms of research outputs, feeling inherent pressure to forward all data channels as useful in a given context (e.g., in terms of predictive value) or present a multimodal approach as better than a unimodal approach without appropriate testing. For instance, a researcher may hesitate to disseminate findings that EDA has negligible predictive value compared to eye tracking, if the researcher has intentions to continue submitting papers

centered on EDA results and may instead present only data that supports the multimodal approach. Additionally, there are still many methodological questions around the generalizability of multimodal approaches and their data sampling to subject rates. In what contexts is having 100,000 samples of one individual better or worse than 1 sample from 100,000 subjects? Where should researchers attempt to strike the balance between generalizability and data veracity? In the long term it is pragmatic and prudent the field begins to hone in on specific best practices in multimodal (or unimodal) engagement measurement.

Lastly, the present work has several limitations, including depictions of the IMMSE ongoing task dynamics. Whereas delineated boxes may suggest discrete stages, they likely overlap. We have also not made any explicit assumptions about the ontological order or hierarchy of the various types of engagement which may influence their temporal relationships. For example, task analysis during the forethought phase may continue during performance, and self-reflection may overlap with performance. However, the dynamic, integrated, and contextual aspects of the model aim to highlight those interactions between facets of engagement over time. In general, research heuristics, including those regarding the aforementioned data channels and temporal granularity, are always subject to exceptions. For example, fixations are often of a social nature, and social interactions are often brief. However, in the context of measuring facets of learning engagement, it is often the case that fixations are used to examine engagement with learning content, even during collaborative tasks (Vrzakova et al., 2021). An additional limitation is that this chapter reviewed many unimodal and multimodal approaches, however we make no remarks about which of these approaches are best (due largely to conceptual and definitional differences). Moreover, although this work forwards heuristics for measure selection, we recognize that more work is needed in terms of formal review and empirical testing.

7 Concluding Thoughts

In this chapter, we introduced an expansion of the integrative model of SRL engagement (Li & Lajoie, 2021) to include emotional, behavioral, and agentic facets of engagement, based on the interrelated aspects of student engagement (Reeve, 2012). We then briefly reviewed the current conceptual, theoretical, and methodological approaches to measuring engagement, and showcased how the use of multimodal data for this work has contributed to our understanding of engagement. We extended previous literature by proposing a methodological overview to inform the research study design of future testing and validation of the IMMSE using multimodal data. Our methodological overview identifies how different modalities of measurement contribute to the measurement of engagement as it fluctuates within the different phases of SRL. We concluded our chapter with several recommendations for future research and system design.

For engagement-sensitive AIEd systems to be designed with underlying student models and adaptive scaffolding approaches, it is first imperative that the environments be able to accurately detect and infer fluctuating levels of engagement. As such, this work seeks to encourage the use of a theoretically driven model to indicate what types of data are most appropriately suited for inferences at various phases of SRL. For example, while most work has used behavioral markers as evidence of cognitive engagement (e.g., environment interactions), we show that measures of eye-tracking may be better suited for cognitive inferences (i.e., cognitive band level) while log files are the behavioral manifestations of engagement at the task level (i.e., rational band level). While both are examples of engagement, our model allows for a distinction on the type of engagement.

Measuring multidimensional facets of SRL engagement with multimodal data raise issues related to ethics, privacy, bias, transparency, and responsibility (Giannakos et al., 2022). Our model emphasizes research and training on the ethical implications of multimodal data proliferation into various facets of multimodal data including detecting, measuring, tracking, modeling, and fostering human learning with AI-based intelligent systems. As researchers we should be deeply committed to addressing ethical value conflicts that are widely known to be related to AI-based research and development including agency (consent and control), dignity (respect for persons and information systems), equity (fairness and unbiased processes), privacy (confidentiality, freedom from intrusion and interference), responsibility (of developers, users, and AI systems themselves), and trust (by users of systems and of data returned by systems). Conflicts among these values are represented through a range of practical, technical, and scientific problems including (1) who consents and does not consent to participate in research where multimodal data is critical to understand SRL engagement, (2) how much and which multimodal data is collected and from whom and where, (3) training on how to collect, analyze, interpret multimodal data, and (4) access to methods, tools, and techniques to analyze multimodal ethically and scientifically.

We argue future research testing our model fundamentally prioritizes the value of equity and fairness as a guiding principle in all our research practices, following national and international guidelines for ethical multimodal data collection, especially when considering the design of intelligent learning systems (Sharma & Giannakos, 2020). We believe that interdisciplinary researchers must be required to develop equity-focused habits of mind, which include noticing, decoding, and deconstructing machine bias and algorithmic discrimination (Cukurova et al., 2020). For example, researchers need to develop competency in strategies to mitigate AI's reification of systemic forms of social inequality (e.g., racial biases, prejudices). In addition, there are fundamental questions that may cause additional challenges that still need to be addressed by researchers. For example, what are the tradeoffs between consenting to some but not all possible multimodal data and the impacts on potential bias in data interpretation and inferences. How long should multimodal data be retained and in what forms? How is access and data sharing negotiated and coordinated between and across collaborators and academics and industry partners? How are learners made aware of what data are being collected and given options and

agency (not agentic engagement but just agency as people in the world) to have voice in what is being inferred? How will explainable AI be unbiased if human researchers are using algorithms, computational models, etc., that are inherently biases because they have been developed by humans and in most cases still include the human-in-the-loop? These are some of the major challenges that multimodal data pose that will need to be addressed in order to avoid biases, prejudice, and potential abuse and misuse of multimodal data as technological advances make it easier for the ubiquitous detection, tracking, modeling of multimodal engagement data. This work serves as the base for a guide to the future direction for both researchers and instructional designers to improve the capturing and analyzing of engagement in AIEd systems using multimodal data.

References

Amon, M. J., Vrzakova, H., & D'Mello, S. K. (2019). Beyond dyadic coordination: Multimodal behavioral irregularity in triads predicts facets of collaborative problem solving. *Cognitive Science, 43*(10), e12787.

Amon, M. J., Mattingly, S., Necaise, A., Mark, G., Chawla, N., & D'Mello, S. K. (2022). Flexibility versus routineness in multimodal health indicators: A sensor-based longitudinal in situ study on information workers. *ACM Transactions on Computing for Healthcare, 3*, 1. https://doi.org/10.1145/3514259

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science, 26*(1), 85–112.

Antonietti, A., Colombo, B., & Di Nuzzo, C. (2015). Metacognition in self-regulated multimedia learning: Integrating behavioural, psychophysiological and introspective measures. *Learning, Media and Technology, 40*(2), 187–209.

Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology, 10*(3), 229–240.

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of school psychology, 44*(5), 427–445.

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*(5), 369–386.

Ashwin, T. S., & Guddeti, R. M. R. (2019). Unobtrusive behavioral analysis of students in classroom environments using non-verbal cues. *IEEE Access, 7*, 150693–150709.

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist, 50*(1), 84–94.

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior, 96*, 207–210.

Azevedo, R., Taub, M., & Mudrick, N. (2015). Think-aloud protocol analysis. In M. Spector, C. Kim, T. Johnson, W. Savenye, D. Ifenthaler, & G. Del Rio (Eds.), *The SAGE encyclopedia of educational technology* (pp. 763–766). SAGE.

Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 254–270). Routledge.

Azevedo, R., Mudrick, N. V., Taub, M., & Bradbury, A. E. (2019). Self-regulation in computer-assisted learning systems. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 587–618). Cambridge University Press. https://doi.org/10.1017/9781108235631.024

Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., et al. (2022). Lessons learned and future directions of MetaTutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology, 13*.

Azevedo, R., & Wiedbusch, M. (2023). Theories of metacognition and pedagogy applied to AIED systems. In Handbook of Artificial Intelligence in Education (pp. 45–67). Edward Elgar Publishing.

Baker, R. S., Corbett, A. T., Roll, I., Koedinger, K. R., Aleven, V., Cocea, M., et al. (2013). Modeling and studying gaming the system with educational data mining. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 97–115). Springer.

Baker, R. S., Ocumpaugh, J. L., & Andres, J. M. A. L. (in press). BROMP quantitative field observations: A review. In R. Feldman (Ed.), *Learning science: Theory, research, and practice*. McGraw-Hill.

Bernacki, M. L., Byrnes, J. P., & Cromley, J. G. (2012). The effects of achievement goals and self-regulated learning behaviors on reading comprehension in technology-enhanced learning environments. *Contemporary Educational Psychology, 37*(2), 148–161.

Bidwell, J., & Fuchs, H. (2011). Classroom analytics: Measuring student engagement with automated gaze tracking. *Behavior Research Methods, 49*(113).

Bixler, R., & D'Mello, S. (2016). Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction, 26*(1), 33–68.

Boekaerts, M. (2016). Engagement as an inherent aspect of the learning process. *Learning and Instruction, 43*, 76–83.

Boucheix, J. M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction, 25*, 71–84.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology, 155*, 49–62.

Chapman, C. M., Deane, K. L., Harré, N., Courtney, M. G., & Moore, J. (2017). Engagement and mentor support as drivers of social development in the project K youth development program. *Journal of Youth and Adolescence, 46*(3), 644–655.

Cheon, S. H., Reeve, J., & Ntoumanis, N. (2018). A needs-supportive intervention to help PE teachers enhance students' prosocial behavior and diminish antisocial behavior. *Psychology of Sport and Exercise, 35*, 74–88.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.), *Self processes and development* (pp. 43–77). Lawrence Erlbaum Associates, Inc.

Corno, L., & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist, 18*(2), 88–108.

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive–affective states during learning. *Cognition and Emotion, 22*(5), 777–788.

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology, 51*(5), 1441–1449. https://doi.org/10.1111/bjet.13015

D'Mello, S. K., & Mills, C. S. (2021). Mind wandering during reading: An interdisciplinary and integrative review of psychological, computing, and intervention research and theory. *Language and Linguistics Compass, 15*(4), e12412.

D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education, 16*(1), 3–28.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies, 70*(5), 377–398.

D'Mello, S. K., Dieterle, E., & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) measurement of engagement during learning. *Educational Psychologist, 52*(2), 104–123.

Darvishi, A., Khosravi, H., Sadiq, S., & Weber, B. (2021). Neurophysiological measurements in higher education: A systematic literature review. *International Journal of Artificial Intelligence in Education*, 1–41.

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review, 28*, 425–474.

Dever, D. A., Azevedo, R., Cloude, E. B., & Wiedbusch, M. (2020). The impact of autonomy and types of informational text presentations in game-based environments on learning: Converging multi-channel processes data and learning outcomes. *International Journal of Artificial Intelligence in Education, 30*(4), 581–615.

Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: A review. *Smart Learning. Environments., 6*, 1. https://doi.org/10.1186/s40561-018-0080-z

Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(3), 1–21.

Dubovi, I. (2022). Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology. *Computers & Education, 183*, 104495.

Duchowski, A. (2007). Eye Tracking Techniques. In: Eye Tracking Methodology. Springer, London. https://doi.org/10.1007/978-1-84628-609-4_5

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior, 52*, 338–348.

Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. The MIT Press.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research, 74*(1), 59–109.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). Springer.

Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction, 43*, 1–4.

Fredricks, J., Hofkens, T., & Wang, M. (2019a). Addressing the challenge of measuring student engagement. In K. Renninger & S. Hidi (Eds.), *The Cambridge handbook of motivation and learning* (pp. 689–712). Cambridge University Press. https://doi.org/10.1017/9781316823279.029

Fredricks, J. A., Reschly, A. L., & Christenson, S. L. (2019b). Interventions for student engagement: Overview and state of the field. In J. A. Fredricks, A. L. Reschly, & S. Christenson (Eds.), *Handbook of student engagement interventions* (pp. 1–11). Academic Press. https://doi.org/10.1016/C2016-0-04519-9

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological Psychology, 50*(1), 61–76.

Gao, N., Shao, W., Rahaman, M. S., & Salim, F. D. (2020). n-gage: Predicting in-class emotional, behavioural and cognitive engagement in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4*(3), 1–26.

Gevins, A., & Smith, M. E. (2008). Electroencephalography (EEG) in neuroergonomics. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 15–31). Oxford University Press.

Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K., Ochoa, X., & Hammad, R. (Eds.). (2022). *The multimodal learning analytics handbook*. Springer.

Gobert, J. D., Baker, R. S., & Wixon, M. B. (2015). Operationalizing and detecting disengagement within online science microworlds. *Educational Psychologist, 50*(1), 43–57.

Grafsgaard, J., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. (2013). Automatically recognizing facial expression: Predicting engagement and frustration. In *Proceedings of the international conference on Educational data mining*.

Greene, J. A., Oswald, C. A., & Pomerantz, J. (2015). Predictors of retention and achievement in a massive open online course. *American Educational Research Journal, 52*(5), 925–955.

Greene, B. A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist, 50*(1), 14–30.

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*(4), 203–209.

Haerens, L., Aelterman, N., Vansteenkiste, M., Soenens, B., & Van Petegem, S. (2015). Do perceived autonomy-supportive and controlling teaching relate to physical education students' motivational experiences through unique pathways? Distinguishing between the bright and dark side of motivation. *Psychology of Sport and Exercise, 16*, 26–36.

Harley, J. M., Pekrun, R., Taxer, J. L., & Gross, J. J. (2019). Emotion regulation in achievement situations: An integrated model. *Educational Psychologist, 54*(2), 106–126. https://doi.org/10.1080/00461520.2019.1587297

Hassib, M., Khamis, M., Friedl, S., Schneegass, S., & Alt, F. (2017a). Brainatwork: Logging cognitive engagement and tasks in the workplace using electroencephalography. In *Proceedings of the 16th international conference on mobile and ubiquitous multimedia* (pp. 305–310).

Hassib, M., Schneegass, S., Eiglsperger, P., Henze, N., Schmidt, A., & Alt, F. (2017b). EngageMeter: A system for implicit audience engagement sensing using electroencephalography. In *Proceedings of the 2017 Chi conference on human factors in computing systems* (pp. 5114–5119).

Haubert, A., Walsh, M., Boyd, R., Morris, M., Wiedbusch, M., Krusmark, M., & Gunzelmann, G. (2018). Relationship of event-related potentials to the vigilance decrement. *Frontiers in Psychology, 9*, 237.

Helme, S., & Clarke, D. (2001). Identifying cognitive engagement in the mathematics classroom. *Mathematics Education Research Journal, 13*(2), 133–153.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education, 90*, 36–53.

Huang, J., Yu, C., Wang, Y., Zhao, Y., Liu, S., Mo, C., … & Shi, Y. (2014, April). FOCUS: enhancing children's engagement in reading by using contextual BCI training sessions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1905–1908).

Ireland, M. E., & Henderson, M. D. (2014). Language style matching, engagement, and impasse in negotiations. *Negotiation and Conflict Management Research, 7*(1), 1–16.

Jang, H., Kim, E. J., & Reeve, J. (2016). Why students become more engaged or more disengaged during the semester: A self-determination theory dual-process model. *Learning and Instruction, 43*, 27–38.

Järvelä, S., & Bannert, M. (2021). Temporal and adaptive processes of regulated learning – What can multimodal data tell? *Learning and Instruction, 72*, 101268.

Järvelä, S., Veermans, M., & Leinonen, P. (2008). Investigating student engagement in computer-supported inquiry: A process-oriented analysis. *Social Psychology of Education, 11*(3), 299–322.

Kew, S. N., & Tasir, Z. (2021). Analyzing students' cognitive engagement in e-learning discussion forums through content analysis. *Knowledge Management & E-Learning: An International Journal, 13*(1), 39–57.

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science, 330*(6006), 932–932.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences, 16*(12), 606–617.

Kruger, J. L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research, 7*(5).

Lee, J., Song, H. D., & Hong, A. J. (2019). Exploring factors, and indicators for measuring students' sustainable engagement in e-learning. *Sustainability, 11*(4), 985.

Li, S., & Lajoie, S. P. (2021). Cognitive engagement in self-regulated learning: An integrative model. *European Journal of Psychology of Education*, 1–20.

Li, S., Zheng, J., & Lajoie, S. P. (2020). The relationship between cognitive engagement and students' performance in a simulation-based training environment: An information-processing perspective. *Interactive Learning Environments*, 1–14.

Li, S., Lajoie, S. P., Zheng, J., Wu, H., & Cheng, H. (2021). Automated detection of cognitive engagement to inform the art of staying engaged in problem-solving. *Computers & Education, 163*, 104114.

Lin, F. R., & Kao, C. M. (2018). Mental effort detection using EEG data in E-learning contexts. *Computers & Education, 122*, 63–79.

Mandernach, B. J. (2015). Assessment of student engagement in higher education: A synthesis of literature and assessment tools. *International Journal of Learning, Teaching and Educational Research, 12*(2).

Molenaar, I., de Mooij, S., Azevedo, R., Bannertd, M., Järveläe, S., & Gašević́, D. (2022). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior*, 107540.

Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist, 50*(1), 31–42.

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., & Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology, 21*(4), 388–422.

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Nuamah, J. K., & Seong, Y. (2018). Support vector machine (SVM) classification of cognitive tasks based on electroencephalography (EEG) engagement index. *Brain-Computer Interfaces, 5*(1), 1–12.

Oshima, J., & Hoppe, H. U. (2021). Finding meaning in log-file data. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International handbook of computer-supported collaborative learning* (pp. 569–584). Springer.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422.

Papamitsiou, Z., Pappas, I. O., Sharma, K., & Giannakos, M. N. (2020). Utilizing multimodal data through fsQCA to explain engagement in adaptive learning. *IEEE Transactions on Learning Technologies, 13*(4), 689–703.

Penttinen, M., Anto, E., & Mikkilä-Erdmann, M. (2013). Conceptual change, text comprehension and eye movements during reading. *Research in Science Education, 43*(4), 1407–1434.

Perry, N. E. (2002). Introduction: Using qualitative methods to enrich understandings of self-regulated learning. *Educational Psychologist, 37*(1), 1–3.

Perry, N. E., & Winne, P. H. (2006). Learning from learning kits: gStudy traces of students' self-regulated engagements with computerized content. *Educational Psychology Review, 18*(3), 211–228.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology, 40*(1–2), 187–195.

Posada-Quintero, H. F., & Chon, K. H. (2019). Innovations in electrodermal activity data collection and signal processing: A systematic review. *Sensors, 20*, 1–18.

Ramachandran, A., Huang, C. M., Gartland, E., & Scassellati, B. (2018, February). Thinking aloud with a tutoring robot to enhance learning. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 59–68).

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149–172). Springer.

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology, 105*(3), 579.

Reeve, J., Cheon, S. H., & Jang, H. R. (2019). A teacher-focused intervention to enhance students' classroom engagement. In J. Fredricks, A. L. Reschly, & S. Christenson (Eds.), *Handbook of student engagement interventions* (pp. 87–102). Academic Press.

Renninger, K. A., & Bachrach, J. E. (2015). Studying triggers for interest and engagement using observational methods. *Educational Psychologist, 50*(1), 58–69.

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In A. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3–19). Springer.

Salmela-Aro, K., Moeller, J., Schneider, B., Spicer, J., & Lavonen, J. (2016). Integrating the light and dark sides of student engagement using person-oriented and situation-specific approaches. *Learning and Instruction, 43*, 61–70.

Schunk, D. H., & Greene, J. A. (2017). Historical, contemporary, and future perspectives on self-regulated learning and performance. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1–15). Routledge.

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology, 51*(5), 1450–1484. https://doi.org/10.1111/bjet.12993

Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Filipe, V., & Reis, M. J. (2019). Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. *arXiv:1909.12913*. https://doi.org/10.48550/arXiv.1909.12913

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A predictive model to evaluate students' cognitive engagement in online learning. *Procedia-Social and Behavioral Sciences, 116*, 4844–4853.

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist, 50*(1), 1–13.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493–525.

Taub, M., & Azevedo, R. (2019). How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive SRL processes during learning with an intelligent tutoring system? *International Journal of Artificial Intelligence in Education, 29*(1), 1–28.

Taub, M., Sawyer, R., Smith, A., Rowe, J., Azevedo, R., & Lester, J. (2020). The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment. *Computers & Education, 147*, 103781.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.

Terriault, P., Kozanitis, A., & Farand, P. (2021). Use of electrodermal wristbands to measure students' cognitive engagement in the classroom. In *Proceedings of the Canadian Engineering Education Association (CEEA)*.

van der Graaf, J., Lim, L., Fan, Y., Kilgour, J., Moore, J., Gašević, D., et al. (2022). The dynamics between self-regulated learning and learning outcomes: An exploratory approach and implications. *Metacognition and Learning*, 1–27.

Van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. *International handbook of metacognition and learning technologies*, 143–156.

Veenman, M. V. J., & van Cleef, D. (2019). Measuring metacognitive skills for mathematics: Students' self-reports versus on-line assessment methods. *ZDM Mathematics Education, 51*, 691–701. https://doi.org/10.1007/s11858-018-1006-5

Verdière, K. J., Roy, R. N., & Dehais, F. (2018). Detecting pilot's engagement using fNIRS connectivity features in an automated vs. manual landing scenario. *Frontiers in human neuroscience, 12*, 6.

Vongkulluksn, V. W., Lu, L., Nelson, M. J., & Xie, K. (2022). Cognitive engagement with technology scale: A validation study. *Educational Rechnology Research and Development, 70*, 1–27.

Vrzakova, H., Amon, M. J., & D'Mello, S. K. (2021). Looking for a deal! Visual social attention during negotiations via mixed media videoconferencing. *Proceedings of the Association for Computing Machinery: Computer Supported Cooperative Work (CSCW), 4*, 1–35. https://doi.org/10.1145/3434169

Wang, M. T., & Fredricks, J. A. (2014). The reciprocal links between school engagement, youth problem behaviors, and school dropout during adolescence. *Child Development, 85*(2), 722–737.

Wang, Y., Kotha, A., Hong, P. H., & Qiu, M. (2020, August). Automated student engagement monitoring and evaluation during learning in the wild. In *2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom)* (pp. 270–275). IEEE.

West, R. L., & MacDougall, K. (2014). The macro-architecture hypothesis: Modifying Newell's system levels to include macro-cognition. *Biologically Inspired Cognitive Architectures, 8*, 140–149.

Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing, 5*(1), 86–98.

Wiedbusch, M. D., & Azevedo, R. (2020). Modeling metacomprehension monitoring accuracy with eye gaze on informational content in a multimedia learning environment. In *ACM symposium on eye tracking research and applications* (pp. 1–9).

Winne, P. H., Jamieson-Noel, D., & Muis, K. (2002). Methodological issues and advances in researching tactics, strategies, and self-regulated learning. Advances in motivation and achievement. *New Directions in Measures and Methods, 12*, 121–155.

Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of educational psychology, 96*(2), 236.

Wolters, C. A., & Taylor, D. J. (2012). A self-regulated learning perspective on student engagement. In A. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 635–651). Springer.

Xie, K., Heddy, B. C., & Vongkulluksn, V. W. (2019). Examining engagement in context using experience-sampling method with mobile technology. *Contemporary Educational Psychology, 59*, 101788.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91.