# Chapter 5
# Reconfiguring Measures of Motivational Constructs Using State-Revealing Trace Data

**Heeryung Choi, Philip H. Winne, and Christopher Brooks**

**Abstract** This chapter examines opportunities afforded by trace data to capture dynamically changing latent states and trajectories spanning states in self-regulated learning (SRL). We catalog and analyze major challenges in temporally investigating SRL constructs related to a prominent motivational factor, achievement goals. The dynamics of potentially frequent state changes throughout a learning session and across sessions are poorly reflected by self-report survey items typically administered before and after a session or, less informatively, at the beginning of an academic term. Trace data, carefully operationalized, offer substantial benefits compensating for shortcomings of comparatively static survey data. We summarize three recent studies addressing these challenges and characterize learning analytics designed to promote SRL and motivation formed from unobtrusive traces. This approach provides a practical and continuously updatable account of SRL constructs, varying dynamically within and across study sessions. We conclude by proposing a research agenda for learning analytics focusing on guiding and supporting SRL.

**Keywords** Trace data · Motivations · Achievement goals · Self-regulated learning · Dynamic SRL constructs

H. Choi (✉)
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: heeryung@mit.edu

P. H. Winne
Simon Fraser University, Burnaby, BC, Canada
e-mail: winne@sfu.ca

C. Brooks
University of Michigan, Ann Arbor, MI, USA
e-mail: brooksch@umich.edu



V. Kovanovic et al. (eds.), *Unobtrusive Observations of Learning in Digital Environments*, Advances in Analytics for Learning and Teaching,
https://doi.org/10.1007/978-3-031-30992-2_5

# 1 Introduction: Self-Regulated Learning

Improvements in educational technologies have allowed researchers to integrate more unobtrusive trace data into their studies. Trace data are clickstream or log records designed to represent a specified theoretical construct revealed as learners operate on information while learning (Winne, 2020a, b). Trace data are particularly useful in measuring and researching dynamic properties of self-regulated learning (SRL).

Winne models how dynamic SRL states arise using the COPES model: Conditions, Operations, Products, Evaluations, and Standards (Winne, 1997, 2022). Self-regulating learners first identify internal and external *conditions* they perceive can affect tasks. Based on their understanding of those conditions, learners choose and carry out metacognitive and cognitive *operations*, generating *products* as a result. Learners then *evaluate* those products, including experiences arising from operations, gauging their properties using *standards*. For example, suppose a learner is preparing for an upcoming quiz in an Earth Science class. First, the learner considers factors such as knowledge about related topics, effort likely required, and incentives for earning a high grade. They recall difficulty listing the names of planets in the solar system and forecast if it is important to remember those to receive a satisfactory grade. Based on this understanding, they design a mnemonic device to assemble the names of planets in the solar system in serial order from the sun outward. After applying the mnemonic, they evaluate its utility using standards such as confidence they will be able to recall all the planets' names in the correct order and effort to encode this information.

SRL is recursive. While working on a given task, a learner could operationalize several COPES learning events to unfold SRL across the learning session and beyond. For example, after a cycle of SRL ends in the evaluation state, a learner might be highly satisfied with their product, such as the invented mnemonic device. This evaluation result could affect an upcoming SRL cycle; motivation might increase since they predict they could easily apply this same tactic to other cases – a high efficacy expectation – and review all the materials more quickly than they expected. Both have high incentive. That is, SRL is a dynamic progression of COPES learning events emerging and contingently unfolding throughout a task, a semester, or an academic year.

# 2 Dynamic Nature of Motivation

## *2.1 How to Capture Motivation*

Motivation is an internal cognitive state that provides reasons for choices learners make about behavior (Kleinginna & Kleinginna, 1981; Winne & Marzouk, 2019). Motivation is often measured by asking learners to rate a motivational construct, such as achievement goals, or by time spent on tasks (Ames & Archer, 1988; Ames, 1992; Elliott & Dweck, 1988; Masgoret & Gardner, 2003). In the COPES model, motivation is a *condition* that contributes to (1) learners' plans, (2) choices about how to approach a task, and (3) forecasts about how to adapt to *operations* to improve work as tasks unfold. Motivation also plays an important role in the *evaluation* facet of COPES; motivation provides reasons for selecting *standards* to judge incentives associated with *products*.

Achievement goal theory generally explains goals using two dimensions: (1) mastery-performance and (2) approach-avoidance. The mastery-performance dimension differentiates the product learners pursue. It contrasts internal standards, such as joy and satisfaction for learning (i.e., mastery), vs. external standards, such as letter grades or ranking and performance with respect to peers. The approach-avoidance dimension contrasts whether learners: (1) seek to acquire desired stimuli (approach) or evade undesired stimuli (avoidance) (Ames, 1992; Nicholls, 1984).

As with other SRL constructs, motivation for achievement goals can dynamically change throughout a task and between tasks. For example, Muis and Edwards (2009) investigated goal changes between similar and different tasks. In both circumstances, they found evidence for both *goal switching*, replacing one goal with another, and *goal intensification*, increasing one's endorsement of an initial goal. They also found mastery-approach goals and performance-avoidance goals were less stable than performance-approach goals, a finding aligned to a previous study (Fryer & Elliot, 2007). Tuominen-Soini et al.'s (2011) studies also showed the dynamic nature of motivation, and they detected changes in Finnish students' achievement goals both between and within a school year. Using latent profile analysis of survey data to develop individual learners' motivational profiles, approximately 35% of students modulated their motivational profiles to reflect similar goal profiles while 5% of students completely changed their goals.

Considering the recursive nature of SRL, goal changes should be expected as learners traverse states in their work. Learners' initial goals may be formed using incomplete information about *conditions*, such as task difficulty. After some time, learners may update goals if *products* generated based on their incomplete understanding of conditions lead to an unsatisfactory evaluation relative to *standards*. In Fryer and Elliot's work (2007), substantial goal changes were more frequent after an initial task than after subsequent tasks, which showed how learners acquired more information from the initial encounter with a task and adjust their goals accordingly in the subsequent tasks.

When goals change, other COPES states may change accordingly, as learners may deem it useful to revise *operations*, hence affecting *products*. New *standards* for *evaluations* may also be adopted. For example, after trying a new strategy to solve the previously attempted math problems and evaluating the new strategy as successful, a learner may perceive greater efficacy for problem-solving and choose to attempt slightly more challenging "extra points" exercises. This goal might change again depending on the pace of work on those more challenging problems, with concomitant changes in self-efficacy depending on whether pace is evaluated as "fast" or "slow."

To more fully understand why motivation changes, and to predict more accurately if and how it will change, it is important to develop fuller accounts of the contexts in which change is observed. Without such contextual information, we suggest it will remain difficult to understand and assess learners' goal changes and design potential improvements to learning experiences.

## *2.2 A Role for Trace Data in Motivational Studies*

Collecting contextual information about motivation and its roles in dynamic SRL is likely to be more informative and authentic if researchers adopt methods to gather fine-grained and unobtrusive data. Log and clickstream data, not yet common in motivational research, offer potential to sharpen research in two ways. First, if unobtrusively generated, such data can be gathered across the timeline of tasks with minimal to no interference. This affords detecting goal changes as they materialize in context. Second, because more detailed information can be captured about external conditions both preceding and at the point of change in motivational states, trace data set a stage for theorizing more productively about how to support learners' motivation while at the same time developing fuller pictures about how SRL relates to developing achievement outcomes. In this effort, it is important to engineer data gathering methods that minimize intrusions on and distortions to learners' authentic learning experiences.

In contrast to the potential benefits of online trace data to reflect changing conditions and motivational dynamics, motivation has been mostly measured using self-report measures – surveys and questionnaires – which often are not sufficiently fine-grained and task-specific (Winne, 2020a, b). The scope of that methodology is broad, ranging across studies and domains from sports to psychology to a residential mathematics course for K-12 learners to distance learning for adult learners (Elliot & Murayama, 2008; Jang & Liu, 2012; Luo et al., 2011; Remedios & Richardson, 2013; Wolters, 2004; Seijts et al., 2004; Chen et al., 2012; Gutman, 2006; Botsas & Padeliadu, 2003; Beck & Schmidt, 2013; Dickhäuser et al., 2021; Janke & Dickhäuser, 2019; Giota & Bergh, 2021).

One challenge to validly interpreting survey responses is that they usually ask learners to aggregate learning experiences across multiple contexts (Turner & Patrick, 2008; Winne, 2010). For example, survey questions often include phrases such as "generally" or "during an exam." This prime is intended to ensure that learners' responses to varying contexts would be consistent for that context. This may not be the case, especially when learners actively self-regulate approaches to learning, as illustrated by research previously cited (see also Hadwin et al., 2001). Moreover, it is usually impractical to administer surveys frequently enough to collect fine-grained data tracking goal changes across the timeline of a single task. When the same or similar questionnaires are given every day, we predict learners acclimate to reporting a generalized "mean experience" rather than taking careful account of varying conditions, particularly if the setting provided in the survey's instructions is not tailored to each administration. While think-aloud or interview methods might lessen this hazard, those methods face other challenges. For example, both for surveys and interviews, learners might respond not based on actual actions but on their knowledge or expectation about which action is recommended for effective learning (Pintrich, 2000). Thus, both accuracy of memory and responses to survey items are in question.

Another issue besetting self-report measures in motivational studies is that learners might not be fully attentive to or willing to report changes in motivational states. Learners' decisions to change goals might be habitual (automated cognition) to the extent that motivation changes go unnoticed. In such instances, fleeting goal changes within a task could be missed in self-reported data. Trace data may be able to complement self-report data in ways that lessen this source of unreliability.

## 3 Critiques of Recent Studies

In this section, we review three recent studies investigating motivation, each of which collected trace data. We reflect on current methodologies and analyze them to suggest directions for future research. We searched the literature using Google Scholar with three queries: "trace data motivation," "log data motivation," and "clickstream motivation." We then chose reports (1) written in English, (2) using unobtrusive trace data to measure learners' motivation, and (3) published in refereed international conferences or refereed journals. Only a few studies could be identified, all showing commonalities in approaches and making approximately equivalent recommendations for future research. We selected three representative studies for analysis here.

Each study was reviewed using a common schema: theoretical framework, contexts, data and indicators, and data analysis and results. The theoretical framework reveals how tightly the approach in each study connects to motivational theories. Approaches include overall study design, operational definitions of indicators, and interpretations of results in relation to theoretical support. Contexts describe where data were collected. Data and indicators examine types of data collected and which indicators were generated from data. Finally, data analysis and results analyze what and how the findings of each study were drawn.

### *3.1 Hershkovitz and Nachmias (2008)*

The main goal of Hershkovitz and Nachmias' (2008) study was to build a conceptual framework to measure motivation using log data.

### 3.1.1 Theoretical Framework

From previous literature, the researchers defined three dimensions of motivation: engagement (how strong motivation is), energization (how long motivation is maintained and direction of motivation), and source (if motivation is internal or external). The framework was used to identify indicators relevant to each of these motivational dimensions.

### 3.1.2 Contexts

Data were collected in a self-paced online system teaching Hebrew vocabulary. Learners could mark each word or phrase as “well known,” “not well known,” or “unknown” based on familiarity. There were five instructional choices the system offered to learners: memorizing where learners could see words and their meanings, practicing where learners see words without meanings, searching for specific words, gaming, and self-testing.

### 3.1.3 Data and Indicators

Data analyzed were secondary; that is, the analyses were conducted on pre-existing data not collected specifically for the study. Each row of these secondary log data recorded a session of a learner’s activity studying vocabulary. A session was initiated when a learner entered the system and ended when a learner closed the system window. Each log also included attributes such as the start and stop timestamp for each session, the number of words that learners marked as “known,” and other actions carried out in the system.

After inspecting raw data for a small number of cases ($N = 5$), the authors identified seven potential indicators of motivation: proportion of time on task, average session duration, pace of actions performed, proportion of words for which learners changed their judgment of familiarity while studying, average time between sessions, proportion of examination events, and proportion of game events.

### 3.1.4 Data Analysis and Results

Indicators were examined in a larger dataset ($N = 1444$) and reduced by an undescribed method to a final dataset ($N = 674$). Hierarchical clustering was applied, and clusters were mapped based on indicators of the three dimensions of motivation: engagement, energization, and source, defined by the researchers as mentioned above.

## *3.2 Cocea and Weibelzahl (2011)*

These researchers aimed to identify behavioral patterns indicating online learners' motivation levels from log files and ultimately to support low-motivated online learners.

### 3.2.1 Theoretical Framework

While the researchers reviewed several prior studies investigating particular motivational states such as confidence and effort, their study adopted a general view of motivation as engagement in learning activities. Further distinctions were not a focus in this research.

### 3.2.2 Contexts

Log data were collected from *HTML -Tutor*, a free online introductory course on HTML. The researchers described the course as interactive but did not provide further details about types of materials or tools, e.g., lecture videos, discussion forums, and the interactive code editor learners used. The amount of material in modules was not described.

### 3.2.3 Data and Indicators

Timestamped data logged for this research were secondary. Data included events such as login, logout, page access, clicking a hyperlink, using a glossary feature, and searching. From raw log data, the authors created indicators for each participant in the study – performance on tests, time spent reading, number of accessed pages, and time spent on tests – which they used to predict learners' motivation levels. The authors also created a binary indicator of motivation level using rules they established. For example, spending at least 60 s per page on average was considered engaged, while spending less than 20 s per page was categorized as disengaged. This motivation indicator was used to label each learner's overall engagement level as engaged or disengaged. The authors did not report the volume of log data they used to create indicators.

### 3.2.4 Data Analysis and Results

Log data files of 20 learners were analyzed using the Waikato Environment for Knowledge Analysis (WEKA) system (Witten et al., 1999). In this data analysis, the four indicators except for motivational level were entered into decision trees to predict learners' motivation levels. The motivation level indicator was used as a gold standard to evaluate decision tree predictions. Analysis classified learners as engaged if they spent more than 45 min on reading and showed a performance either above 63% or below 49%.

The authors attributed relatively lower engagement for learners with medium-level performance (between 49% and 63%) to the learners' confidence. The authors interpreted these learners did not invest much effort to improving their performance because the learners judged their level of achievement was good enough.

## *3.3 Zhou and Winne (2012)*

The aim of this study was to examine potential differences in achievement goals measured by self-reported surveys and by log data.

### 3.3.1 Theoretical Framework

Goal orientation theory was the main theoretical framework adopted in this research. The authors criticized self-report measures used in prior research as too divergent in operationalizations of goal orientations. They also questioned whether respondents validly reported goal orientations because self-report items were framed at too coarse a grain size. These researchers designed log data and indicators to capture four goal orientations: mastery-approach, mastery-avoidance, performance-approach, and performance-avoidance. They examined the predictive power of traces of goal orientation as compared to self-report data.

### 3.3.2 Contexts

Zhou and Winne's study generated primary data in a one-hour-long lab experiment. Learners first responded to the Achievement Goal Questionnaire (Elliot & McGregor, 2001) then read an article about hypnosis. After studying, they took achievement tests posing five multiple-choice items and five short-answer questions.

Two measures of goal orientation were obtained: the self-report Achievement Goal Questionnaire (Elliot & McGregor, 2001) and trace data generated as learners studied in a software system, gStudy. gStudy was a Chrome extension that provided tags and hyperlinks to learners allowing them to choose sources of help to prepare for the achievement test. Each tag and each hyperlink was mapped to one of the four goal orientations according to their labels (e.g., tag: "Reread to avoid misinterpretation" tracing mastery avoidance). Tagging and clicking hyperlinks traced expressions of goal orientations while studying.

#### 3.3.3 Data and Indicators

Zhou and Winne's raw trace dataset was composed of learners' clicks on hyperlinks and tags applied. Counts of traces were used to form four behavioral indicators, one for each facet of goal orientation. For example, if a learner created a tag representing mastery-approach goal orientation five times, that count divided by the total number of all goal orientation traces formed the indicator of mastery-approach goal orientation.

#### 3.3.4 Data Analysis and Results

Results showed correlations between self-reports and trace data were not statistically detectable ($p \geq .05$). Their blocked multiple regression analyses revealed trace-based indicators were statistically better predictors of learners' achievement than any survey-based indicators ($p \leq .01$). Furthermore, all trace-based indicators except one for mastery-avoidance orientation showed a strong Kendall's *tau b* coefficients predicting achievement ($p \leq .01$). None of the survey-based indicators, on the other hand, were a statistically detectable predictor of achievement ($p \geq .05$).

### *3.4 Critiques of the Select Studies*

#### 3.4.1 Importance of Design Processes

Trace data may be noisy, i.e., contaminated with sources of variance not relevant to target constructs. Thus, one important task for researchers is identifying and minimizing noise to enhance the resolution of trace data (Krumm et al., 2022; Winne, 2014). For example, a clickstream datum showing a learner clicked a hint button could indicate various motivation constructs, e.g., simply exploring a software feature vs. attempting to overcome difficulty vs. gaming the system. Noise contaminating trace data, as with any kind of data, jeopardizes valid interpretation. Carefully designing trace data collection in consideration of theories, contexts, and research questions is essential.

In two research cases we reviewed, data were secondary, and the design rationales for motivational indicators were minimally explained. This severely challenges the validity of drawing correspondences between trace data and constructs each trace it intended to represent. Hershkovitz and Nachmias (2008) used secondary data and did not justify how those data represent learners' motivation. They also mentioned they chose indicators used in previous work, but information was minimal about operational definitions as explicit expressions of theory. For instance, their indicator *timeOnTaskPC*, the total time of active sessions divided by the total time logged, was presented as a measure of the engagement dimension of motivation. Because the time learners are logged in can be spent on many different activities, e.g., exploring features of the interface or responding to text messages received on a smartphone, we suggest time on task metrics are typically overly broad and imprecise indicators of motivation.

Similarly, Cocea and Weibelzahl's (2011) use of secondary data prevented designing traces that more directly represent motivational constructs. Furthermore, insufficient explanation regarding their design process limits interpretations of their results. They provided only a table of indicator names and general indicator descriptions. For example, an indicator *NoPages* was described as the number of accessed pages. That indicator is potentially unrepresentative of motivation if a website's architecture requires learners to pass through landing pages or where one website provides a single scrolling page while that same volume of information at another website is distributed across separate pages linked by a "Next" button. Also, it is unclear whether a learner's retreat to a previously viewed page is included in the count *NoPages*. Retreat may be a strong indicator of a learner's motivation to reinstate forgotten information or to monitor clarity about previously studied content.

In contrast, Zhou and Winne (2012) detailed theoretical grounding for designing indicators in their study. While their descriptions might have been more detailed, traces they logged about learners' behavior are explicitly mapped to specific aspects of achievement goal orientation theory according to Elliot and McGregor (2001). This approach permits constructive critique about how those operational definitions manage noise and introduce subjectivity in traces vis à vis constructs they are designed to indicate.

None of these three studies considered motivation change within a learning session even though previous studies show motivation is dynamic (Senko and Harackiewicz 2005). In Hershkovitz and Nachmias's study (2008), the duration of each learner's interaction with the learning platform ranged from 3 weeks to 3 months. Zhou and Winne's (2012) study was just an hour-long and its context was a lab study. Cocea and Weibelzahl (2011) did not clarify how long learners' interaction with *HTML-Tutor* lasted. Methodologies designed to take account of motivational dynamics across the timeline of learner's engagement would be more revealing.

### 3.4.2 Weak Evaluation Process of Indicators

After generating indicators to measure constructs, it is important to evaluate them in the context of a specific study for future researchers. Two of the three studies we reviewed did not describe an evaluation process: Hershkovitz and Nachmias (2008) did not evaluate their indicators, and Cocea and Weibelzahl (2011) evaluated their indicators by comparing classification results against hand-labeled data identifying whether a learner was engaged or disengaged. While Cocea and Weibelzahl's approach is a step in the right direction, there is no outside criterion beyond the researchers' judgment. As well, some decisions could be considered arbitrary, e.g., choosing "less than 20 seconds spent per page" as the standard for disengagement instead of 15 or 25 s. They chose this threshold based on estimated times for reading a page or working on a test without explaining how these times were estimated. Without sharing further contextual information, such as how many tasks learners had available to work on and some metric of required "steps" to complete each task, it is hard to evaluate the likelihood that indicators of engagement usefully reflect learners' motivation.

Zhou and Winne (2012) evaluated trace-based goal orientation in two ways. First, they examined the correspondence between goal orientations measured by their trace-based indicators and a widely used self-report measure. When they observed weak correspondence between trace and self-report indicators of goal orientation, they examined posttest performance to identify which indicator more strongly aligned to theory's predictions of achievement. They concluded their trace indicators outperformed self-reports as indicators of goal orientation in their study context.

### 3.4.3 Lack of Discussion on How Trace Measures Were Introduced to Users

While trace data can represent learners' dynamic motivation unobtrusively and, arguably, more directly than self-reported data, benefits may be undermined if the user experience which creates the trace measures is not carefully considered. Traces inherently require the learner to engage with content, e.g., highlight it, or use features in an interface, e.g., a menu of options or a button, controlling software features. If learners are unaware those kinds of engagements are available or do not understand how a software feature functions, trace data will not be generated regardless of learners' motivation, cognition, or metacognition. If the method for creating a trace is perceived to be overly effortful, requires complex maneuvers in the learning environment, or slows the pace of a learner's work too much, learners will avoid the feature that generates trace data. Learners are generally uninterested (and unaware) of the trace data being created, so features of the environment which are instrumented for trace data must provide a clear benefit to the learner in order to be used.

In other words, designing tools to gather trace data requires careful attention to the user experience. In some cases, it may be necessary to provide initial training to learners about how to use trace-generating tools to ensure they understand and appreciate how the tool can be useful in learning. Where the tool appears to learners as a socially desirable property or can be used excessively to game the system, further cautions apply to designing it. We suggest a general guideline: Learning tools which have been designed with tracing methods must have perceived utility to the learners.

Two of the three studies we reviewed did not address the issue of how trace data were related to learner motivational states. For example, in Hershkovitz and Nachmias' (2008) online system teaching Hebrew vocabulary, learners could mark each word or phrase depending on their confidence. Furthermore, learners could use other features such as searching, memorizing, and self-testing. Yet, it is unknown how obvious these features were to learners or how well they were integrated into purposes of the learning task. If data showed learners did not use features after a few attempts or only a few learners continued to use these features, questions arise about the extent to which traces measure enough of behavior and kinds of behavior that serve research goals.

## 4 Proposals

What does our review of three representative studies suggest for improving online measures of motivation in research and contributing to advancing motivation theories?

### *4.1 Implementing Design Framework*

There are only a few examples of unobtrusive trace indicators of motivation in the field of learning analytics. Researchers aiming to represent motivation using trace data appear likely to design novel indicators rather than build on prior work where strengths and weaknesses of indicators and data designs can be assessed in particular contexts. Thus, we recommend it is important to meticulously inspect each study's design to analyze how and the extent to which it reduces noise and explicitly details key features of the method for generating and logging trace data.

One approach may be using a structured design framework such as the Evidence-Centered Design (ECD) (Mislevy & Steinberg, 2003; Mislevy & Haertel, 2007). ECD is a framework that evaluates assessments designed to permit learners to display knowledge or skills. In this approach, assessment is broadly considered as an argument to be supported by evidence describing learners' latent constructs, such as motivation, knowledge, or a particular skill. Ideally, it would be possible to reliably and validly ascribe a motivational state based on low-noise instances of behavior and patterns.

In particular, ECD's design pattern (Gamma et al., 1995; Alexander et al., 1977) helps researchers build a more solid rationale for their designs of indicators. Implementing a design pattern is often approached by completing a table identifying attributes of a construct and their operational definitions in particular study contexts. For example, suppose a researcher aims to distinguish learners' achievement goals focusing on earning higher final grades (i.e., performance-oriented goals) from mastering learning materials for satisfaction (i.e., mastery-oriented goals). To measure the performance-oriented goal, the researcher designs an indicator as follows: If a learner clicks a hyperlink "critical concepts for the final exam," that could supply evidence of performance-orientation. While implementing the design pattern, the researcher should explain the rationale detailing how this indicator could be strong evidence for performance-oriented goals. Furthermore, the researcher should consider what alternative latent constructs this indicator might represent. For example, learners may click a link simply out of curiosity, not because they hold performance-oriented goals. Through this careful process, a researcher could thoroughly inspect a rationale for a proposed indicator design, potentially improving the design for generating data in ways that improve validity when interpreting data.

Beyond dutiful attention to principles of ECD and considerations Winne (2020a, b) forwarded to improve validity of inferences made and actions (subsequent instructional interventions) based on trace data, we recommend four characteristics for indicators.

First, it should be almost intuitively obvious to learners that information they generate using a tool has value for learning. Highlighted information, for example, eases burdens of locating content judged as meriting review or attention when studying for an examination. Tags greatly facilitate sorting information into categories, e.g., tasks not to be forgotten and major bins in a discipline (e.g., major theorists, disproven hypotheses, useful shortcuts in procedures).

Second, effort required to use a tool should be minimized, thereby reducing extraneous cognitive load. Most undergraduates highlight often and have extensive experience highlighting text in pdf readers or via an extension added to their favorite web browser. Learning how to highlight text once the toolbar icon or keystroke shortcut is introduced is practically one-trial learning. In general, software designs for tools that generate trace data should follow usual guidelines for optimizing the user experience.

Third, the set of tools available to learners should span options for operating on different kinds of information using different operations that achieve different purposes. Without choice, learners are constrained to display variance in their behavior and corresponding inferred underlying processes that comprise SRL. For example, tools for planning steps in a large task and monitoring progress serve quite different purposes than tools for tagging interesting information worth researching further than tools for re-searching information falling into categories.

Fourth, we conjecture learners may be more inclined to "give a tool a chance" if they are provided reasons the tool is designed the way it is. Having and providing a rationale warranting when and why to use a tool may increase chances learners will trial it.

## *4.2 Evaluating Indicator Designs for Future Studies*

To replicate or adopt suggested indicator designs in future studies, it is important to analyze indicators in particular contexts. Construct validity is the degree to which an interpretation of an indicator is justified regarding the presence or degree of a construct. Construct validity is a key concern when evaluating indicators. External validity refers to the degree to which an indicator can be justifiably interpreted in relation to other variables (Messick, 1987). For example, if previous work generally agrees performance goals and academic performance are positively correlated, an indicator designed to capture performance goals should also have a relatively large positive correlation with performance measurements such as posttest scores.

Among studies we reviewed, only Zhou and Winne (2012) correlated trace indicator data purportedly representing achievement goals with posttest scores. That move helps consolidate not only the validity of their indicator designs but also their study's implications. In contrast, Hershkovitz and Nachmias (2008) and Cocea and Weibelzahl (2011) did not pursue these lines of analysis. Combined with a weaker design framework for their indicators, this omission increases uncertainty about the appropriateness of indicator designs in these two studies as a basis for designing future research.

Accumulating evaluation results in diverse contexts is also essential when attempting to generalize motivational indicator designs based on operational definitions of unobtrusive trace data. After particular indicators have been validated as reliably and informatively capturing specific features of learners' motivation in one specific context, those indicators should be examined in related contexts. This would allow researchers to explore for contextual factors affecting the validity attributed to an indicator design.

Researchers should adapt indicator designs to unfolding and varying conditions, both internal and external to the learner. For example, following success on a timed practice quiz problem, learners might be more motivated to choose more difficult problems when they login to the next study session. This motivational change may well affect goals set, tactics chosen, time allocated, and emotional stance. To more accurately capture and analyze such contextual changes implies adapting indicator designs to reflect factors such as changed difficulty levels and new learning tactics. In the abstract, trace data can detect such fine-grained changes but only when researchers forecast changes that may arise and consider how indicators should be re-designed under those changed conditions.

## *4.3 Introducing Interventions Less Obtrusively*

One potential step to reduce noise in laboratory studies is giving learners time to explore and practice using a given system. For example, Zhou and Winne (2012) provided participants with a short practice session, an important opportunity since they asked learners to use unconventional hyperlinks and tags to trace achievement goals. Although brief, the practice session likely increased the chance learners would use these features.

We also suggest researchers consider carefully the context of trace data before including it in an analysis if it was not designed specifically to measure the constructs of interest. While this suggestion does not mean that secondary trace data cannot be used to support analyses, it is important for researchers to consider how that data was created by the system in response to the context of learning. Misaligned data may lead to inappropriate conclusions about leaner motivational states.

Furthermore, we encourage researchers to design features for generating trace data with considerations for learning contexts. Theoretically elegant tools may generate more noise than signal if not tightly articulated to learning objectives and learners' understanding of purposes. For example, a tool learners can use to tag content *research this* generates a clear picture about learners' intentions to engage with additional content. But what is the motive underlying that plan – curiosity, performance orientation (to find material resulting in a higher score on a research paper), anxiety (that important content will be omitted for a research paper)? Steps to usefully constrain interpretations of those trace data, perhaps by changing the label for the tag, may be elusive but necessary.

## 5 Conclusion

Compared to widely used self-report measures, fine-grained and unobtrusive trace data may often offer stronger alignment to dynamic motivational constructs. Yet, capturing motivational events through trace data remains relatively underexplored in learning analytics, especially how dynamics can be represented to learners and leveraged to guide SRL. Among the few existing studies, rationales for designing indicators of motivation often appear to be insufficiently justified, if at all. This slows advances to theory and curtails the potency of practical recommendations. Our proposals for improving design and validation of indicators that trace constructs should nurture a more rigorous approach to research and the development of more serviceable learning analytics.

## References

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A pattern language: Towns, buildings, construction*. Oxford University Press.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261. http://psycnet.apa.org/journals/edu/84/3/261.html?uid=1993-03487-001

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260–267. https://doi.org/10.1037/0022-0663.80.3.260

Beck, J. W., & Schmidt, A. M. (2013). State-level goal orientations as mediators of the relationship between time pressure and performance: A longitudinal study. *The Journal of Applied Psychology, 98*(2), 354–363. https://doi.org/10.1037/a0031145

Botsas, G., & Padeliadu, S. (2003). Goal orientation and reading comprehension strategy use among students with and without reading difficulties. *International Journal of Educational Research, 39*(4), 477–495. https://doi.org/10.1016/j.ijer.2004.06.010

Chen, Z.-H., Liao, C. C. Y., Cheng, H. N. H., Yeh, C. Y. C., & Chan, T.-W. (2012). Influence of game quests on pupils' enjoyment and goal-pursuing in math learning. *Journal of Educational Technology & Society, 15*(2), 317–327. https://www.jstor.org/stable/pdf/jeductechsoci.15.2.317.pdf

Cocea, M., & Weibelzahl, S. (2011). Can log files analysis estimate learners' level of motivation? *Proceedings of the Workshop Week Lernen - Wissensentdeckung - Adaptivitat, Hildesheim*, 32–35.

Dickhäuser, O., Janke, S., Daumiller, M., & Dresel, M. (2021). Motivational school climate and teachers' achievement goal orientations: A hierarchical approach. *The British Journal of Educational Psychology, 91*(1), 391–408. https://doi.org/10.1111/bjep.12370

Elliot, A. J., & McGregor, H. A. (2001). A 2× 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501. http://doi.apa.org/journals/psp/80/3/501.html

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613. http://doi.apa.org/journals/edu/100/3/613.html

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(1), 5–12. https://doi.org/10.1037//0022-3514.54.1.5

Fryer, J. W., & Elliot, A. J. (2007). Stability and change in achievement goals. *Journal of Educational Psychology, 99*(4), 700–714. https://doi.org/10.1037/0022-0663.99.4.700

Gamma, E., Helm, R., Johnson, R., Johnson, R. E., & Vlissides, J. (1995). *Design patterns: Elements of reusable object-oriented software*. Pearson Deutschland GmbH. https://play.google.com/store/books/details?id=tmNNfSkfTlcC

Giota, J., & Bergh, D. (2021). Adolescent academic, social and future achievement goal orientations: Implications for achievement by gender and parental education. *Scandinavian Journal of Educational Research, 65*(5), 831–850. https://doi.org/10.1080/00313831.2020.1755360

Gutman, L. M. (2006). How student and parent goal orientations and classroom goal structures influence the math achievement of African Americans during the high school transition. *Contemporary Educational Psychology, 31*(1), 44–63. https://doi.org/10.1016/j.cedpsych.2005.01.004

Hadwin, A. F., Winne, P. H., Stockley, D. B., Nesbit, J. C., & Woszczyna, C. (2001). Context moderates students' self-reports about how they study. *Journal of Educational Psychology, 93*(3), 477–487. https://doi.org/10.1037/0022-0663.93.3.477

Hershkovitz, A., & Nachmias, R. (2008). *Developing a log-based motivation measuring tool* (pp. 226–233). EDM. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.4775&rep=rep1&type=pdf

Jang, L. Y., & Liu, W. C. (2012). 2 × 2 Achievement goals and achievement emotions: A cluster analysis of students' motivation. *European Journal of Psychology of Education, 27*(1), 59–76. https://doi.org/10.1007/s10212-011-0066-5

Janke, S., & Dickhäuser, O. (2019). A neglected tenet of achievement goal theory: Associations between life aspirations and achievement goal orientations. *Personality and Individual Differences, 142*, 90–99. https://doi.org/10.1016/j.paid.2019.01.038

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5*(4), 345–379. https://doi.org/10.1007/BF00992553

Krumm, A. E., Coulson, A., & Neisler, J. (2022). Defining productive struggle in ST math: Implications for developing indicators of learning behaviors and strategies in digital learning. In *LAK22: 12th international learning analytics and knowledge conference*. https://doi.org/10.1145/3506860.3506901

Luo, W., Paris, S. G., Hogan, D., & Luo, Z. (2011). Do performance goals promote learning? A pattern analysis of Singapore students' achievement goals. *Contemporary Educational Psychology, 36*(2), 165–176. https://doi.org/10.1016/j.cedpsych.2011.02.003

Masgoret, A.-M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning, 53*(S1), 167–210. https://doi.org/10.1111/1467-9922.00227

Messick, S. (1987). Validity. *ETS Research Report Series, 1987*(2), i–208. https://doi.org/10.1002/j.2330-8516.1987.tb00244.x

Mislevy, R. J., & Haertel, G. D. (2007). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6–20. https://doi.org/10.1111/j.1745-3992.2006.00075.x

Mislevy, R. J., & Steinberg, L. S. (2003). Focus article: On the structure of educational assessments. *Research and Perspectives*. https://www.tandfonline.com/doi/abs/10.1207/S15366359MEA0101_02?casa_token=VRG8zQQhzU0AAAAA:r4ZnL1_0WgWOglxNoBevXR4IUC7BxGmdG2JH5gz5hq9Qq_XwrCKfwwhNMHzT1BhKj_s15VNyJeEHbA

Muis, K. R., & Edwards, O. (2009). Examining the stability of achievement goal orientation. *Contemporary Educational Psychology, 34*(4), 265–277. https://doi.org/10.1016/j.cedpsych.2009.06.003

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*(3), 328–346.

Pintrich, P. R. (2000). Chapter 14 – The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic. https://doi.org/10.1016/B978-012109890-2/50043-3

Remedios, R., & Richardson, J. T. E. (2013). Achievement goals in adult learners: Evidence from distance education. *The British Journal of Educational Psychology, 83*(Pt 4), 664–685. https://doi.org/10.1111/bjep.12001

Seijts, G. H., Latham, G. P., Tasa, K., & Latham, B. W. (2004). Goal setting and goal orientation: An integration of two different yet related literatures. *Academy of Management Journal, 47*(2), 227–239. https://doi.org/10.5465/20159574

Senko, C., & Harackiewicz, J. M. (2005). Regulation of achievement goals: The role of competence feedback. *Journal of Educational Psychology, 97*(3), 320–336. https://doi.org/10.1037/0022-0663.97.3.320

Tuominen-Soini, H., Salmela-Aro, K., & Niemivirta, M. (2011). Stability and change in achievement goal orientations: A person-centered approach. *Contemporary Educational Psychology, 36*(2), 82–100. https://doi.org/10.1016/j.cedpsych.2010.08.002

Turner, J. C., & Patrick, H. (2008). How does motivation develop and why does it change? Reframing motivation research. *Educational Psychologist, 43*(3), 119–131. https://doi.org/10.1080/00461520802178441

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology, 89*(3), 397. http://psycnet.apa.org/fulltext/1997-05647-001.html

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267–276. https://doi.org/10.1080/00461520.2010.517150

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning, 9*(2), 229–237. https://doi.org/10.1007/s11409-014-9113-3

Winne, P. H. (2020a). Commentary: A proposed remedy for grievances about self-report methodologies. *Frontline Learning Research*. https://eric.ed.gov/?id=EJ1260776

Winne, P. H. (2020b). Construct and consequential validity for learning analytics based on trace data. *Computers in Human Behavior, 112*, 106457. https://doi.org/10.1016/j.chb.2020.106457

Winne, P. H. (2022). Modeling self-regulated learning as learners doing learning science: How trace data and learning analytics help develop skills for self-regulated learning. *Metacognition and Learning*. https://doi.org/10.1007/s11409-022-09305-y

Winne, P. H., & Marzouk, Z. (2019). Learning strategies and self-regulated learning. In J. Dunlosky (Ed.), *The Cambridge handbook of cognition and education* (Vol. 729, pp. 696–715). Cambridge University Press, xviii. https://doi.org/10.1017/9781108235631.028

Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G., & Cunningham, S. J. (1999). *Weka: Practical machine learning tools and techniques with Java implementations*. https://researchcommons.waikato.ac.nz/handle/10289/1040

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), 236–250. https://doi.org/10.1037/0022-0663.96.2.236

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction, 22*(6), 413–419. https://doi.org/10.1016/j.learninstruc.2012.03.004