



## Discussion

### Paradigmatic Issues in State-of-the-Art Research Using Process Data

Philip H. Winne <sup>a</sup>

<sup>a</sup> Faculty of Education, Simon Fraser University, Canada

#### Abstract

*Learning science is enthusiastically adopting new instruments to gather physiological and other forms of event data to represent mental states and series of them that reflect processes. In an attempt to provoke more thought about this kind of research, I suggest paradigmatic issues relating to data, analyses of them and interpretations of results. I advocate we not label these data as “objective.” Instead, we share a subjective interpretation of them. I argue propositions about validity need more nuance. Bounds on generalization related to so-called ecological validity are rarely empirically justified. When researchers transform raw data before analysis and when analytic methods partition variance, interpretations of results omit key qualifications. I posit emotion and motivation be positioned in theory as moderators rather than mediators because agentic, self-regulating learners make and revise knowledge by choosing forms of cognitive engagement in a context where they interpret arousal. I note that researchers’ anchor interpretations of process data in learners’ accounts. This creates a tautology that troubles usual notions of reliability. Finally, I recommend research involving process data turn more toward helping learners identify conditions of learning that spark arousal so learners can regulate motivation and emotion. This leads to a surprise: Treating learners as individuals and helping them identify triggers of arousal may recommend learning science cast emotions and motivation as epiphenomena.*

*Keywords: validity; trace data; agency; motivation; emotion; self-regulated learning, learning science paradigm*



## Paradigmatic Issues in State-of-the-Art Research Using Process Data

A great deal of recent research has investigated relations between learners' affective states, brain states and other physiologically-related variables to traditional measures of achievement, motivation and upcoming indicators of cognitive processing called traces. Articles in this special issue represent a broad and high-quality sample of these efforts. They boldly explore newer approaches to gathering data, tackle challenges in analyzing data with unconventional properties, and suggest new views of frameworks to account for learning processes that create outcomes.

It is almost certain that no research study is conceptually faultless or methodologically perfect. Interpretations and implications researchers draw about their results arise in that context and, thus, are debatable. In this article, I do make conventional critiques about whether this or that instrument or analytic approach is faulty or whether another is likely more appropriate. Instead, I describe from my perspective paradigmatic issues about these kinds of research. My aim is to provoke thinking not about any particular study but about fundamental characteristics of this up-and-coming line of research.

### 1. Process and Trace Data Are a Step Forward But Not the Truth

One description often applied to process data is they are objective. Depending on what one means by “objective” this is a valid interpretation or it is wrong. It is fine to label process data as objective in the sense that “reasonable” observers can agree whether an event occurred. I prefer to think of this as shared subjectivity rather than objectivity. A second sense of the concept of objectivity is wrong. From this perspective, data are conceptualized as incapable of having any other value. I elaborate.

Data such as a gaze duration or a button click are verifiable – a learner's gaze settled on a particular area of interest for  $k$  or more milliseconds. A button was clicked. The metric is nominal and boundaries are definite. For data like these the question is: What are properties the counting metric? If each gaze period or each click is identical, each event may count as “1” and a total period of gazing at something in particular or the sum of clicks can be identified by adding each instance. In this context where data are labeled “objective” theoretical constructs must be considered.

First, I take as axiomatic – believed without proof – learners are agents. They are in control of what they think about. I fend off an immediate counterclaim about mental activities that might be considered a “cognitive reflex.” Some mental activities are genuine reflexes. The startle response is an example. Genuine reflexes like these are very infrequent in everyday learning situations, so I treat them as rare and genuine anomalies. Other apparent cognitive reflexes are learned cognitive routines that have been automated through extensive experience, particularly practice with feedback. Learners are supposed to develop and automate such routines and use them as they study and collaborate. Understanding others' speech at everyday rates of utterance is an example. Other examples include number facts such as  $4 \times 8 = 32$ , raising a hand to be recognized in discussion, and subvocalizing ROY G BIV to name colors in the visible spectrum in order of their wavelengths from shorter to longer. As well, education encourages students to disassemble



other cognitive routines that are disciplinary or social misconceptions. Examples are: denominators must be identical to multiply fractions, females are not adept at math and maintenance rehearsal is the best tactic to promote recall .

Because learners are agents, observers need more information than gaze duration or clickstream events to validly interpret an observation. Suppose a learner gazing at a particular region in a diagram of an electric circuit could use software to identify that region (e.g., enclose it in an ellipse) and tag it “confusing.” These extra data – the region enclosed plus the tag – signal what the learner was thinking that caused gaze to linger. The learner judged (metacognitively monitored) she was challenged to understand something about that part of the circuit. She was motivated to make a permanent record of that state of mind, so she drew an ellipse. Quite likely, she plans to search later – ask the teacher or a peer, comb the internet – to locate information about this and other content tagged “confusing” to resolve these confusions. Suppose a button in a software tool was labeled “See more ....” A click of that button signals the learner is seeking additional or elaborative information beyond what is presented in the current display. Clicking the button represents interest, an expectation useful material will be found at the resource linked to the button, and an efficacy expectation understanding can be enhanced by accessing that new information.

These examples illustrate trace data (Winne, 1982). The best instances of trace data couple an a verifiable event – gaze lingering for a measured time on a particular bit of information, a button clicked, a tag applied – with a convincing theoretical claim about what that event “means.” When learners generate trace data without having to do much more than they normally would do as they study – when the data are ambient in the sense means used to generate data are integrally involved in ways learners normally engage with information (see Winne, Teng, Chang, Lin, Marzouk, Nesbit, Patzak, Raković, Samadi, & Vytasek, 2019) – observers have the additional information needed to construct well founded inferences. Note, however, such inferences are nonetheless grounded in a theoretical framework.

An event datum or string of event data create an opportunity to ask, Why did that event occur? Why is the string of events shaped as it is? Learning science is keen to notice events and characterize strings of them. But event data are not objective. People notice anomalous events because they are unexpected and interesting with respect to subjective schemas that describe the world. When researchers use instrumentation to notice an event, theories underlie the mechanisms that allow the instrument to record the event. In short, process data and especially trace data are inherently and inescapably subjective. What we mean by the label “objective” is really that we share subjectivity.

## 2. Claims about Validity Often Overreach

Validity is a concept often misinterpreted. As Messick (1989) argued, validity is not a property of an instrument, a protocol or a setting. Settings, instruments and protocols do not *have* validity. Validity is more nuanced. It is a property of an inference or an interpretation. People construct inferences and interpretations. I consider three important cases.



## 2.1 To What Does Validity Apply?

Statements of relationship or causation – the findings reported about research in learning science – have a degree validity. That degree is proportional to the extent constructs named in stating a finding correspond to operational definitions that describe how data were generated. One form of this relationship was just discussed in regard to trace data. Other less obvious cases need address.

### *Transformations of Data Also Transform Constructs*

Suppose a researcher transforms raw data. Two common kinds of transformations populate our research literature. One is transformations of scale, such as a log transformation of continuous scores or an arcsine transformation of proportions. These are often used to reshape a distribution of data so it more closely matches a Gaussian (normal) distribution that is better suited to many inferential statistical methods. A second widely used transformation of data is statistically partitioning and removing variance from a variable. Examples are standardized partial weighting (e.g., regression) coefficients in a linear modeling analysis.

Both scale and variance partitioning transformations of data change raw scores into new scores. These new transformed scores are then analyzed by a statistical or machine learning method. Results of these analytical methods are then framed using words identical to the untransformed data. This is careless phrasing.

The fault is not with the numerical work. It has been rendered appropriate by the transformation. The fault lies in failing to recognize transformed scores introduce an additional operational definition, the transformation. Unlike careful attention to reporting the span of a scale for reporting responses to questionnaire items, the number of options offered in multiple-choice items or sampling rates for physiological processes that vary continuously across time, changes to data wrought by numerical transformations are almost never taken into account.

Why is this important? Numerically transformed data submitted to analytical methods represent a different construct than the construct represented by raw data. But, this difference is not represented in descriptions and interpretations of analytical results. Researchers don't write, "The base 10 log version of our control variable ... ." The nomological network changes when data are transformed (see Winne, 1983). Correlations of raw scores with transformed scores always are less, sometimes much less, than 1.00. Correlations of raw scores with other anchor variables differ, sometimes considerably, from correlations of transformed scores with those anchor variables. Learning science is misled whenever the full operational definition of data is not recognized.

## 2.2 Concerns for Ecological Validity Are Rarely Empirically Justified

A third concern about validity relates to the concept of ecological validity. Variations of this claim are common in the literature of learning science: "Because learners carry out tasks in settings where they learn every day, a study's findings have ecological validity." It follows: "When conditions drift away from those characterizing an everyday setting, e.g., from a regular classroom to a lab, findings lose ecological validity in proportion to the drift."



The error of this claim about lessened generalizability is twofold. First, worry that findings disintegrate in a new setting very rarely have grounding in any or sufficient research demonstrating that factors differentiating the settings actually affect findings. It is too casually presumed without empirical backing that such and such a factor differentiating settings is a cause or moderator of an effect. Two examples are our literature's practice of reporting where a sample originates (e.g., a Western Canadian university) and the proportions of females and males in a sample. To my knowledge, no study has demonstrated geographic locations of Canadian universities moderate variance in findings other than the residential addresses of participants in studies. Or, if one study's participants are 60% female and another study's participants are 48% female, where are studies demonstrating sex of participant is a proven moderator of cognitive processing or emotion? In general, our research traditions have a very sparse catalog of factors that influence the values of population parameters as a function of factors such as location, proportion of females and males, language first spoken, ethnic heritage and the like (see Winne, 2017). In fact, claims about "ecological validity" are more guesswork.

There are straightforward repairs for this fault. Researchers can analyze their data to investigate the extent to which ecological factors moderate findings in their study. To do this places two burdens on research. First, such factors need to be considered at the outset of a study and measured. Second, sample sizes need to be larger. Another solution is to scour the literature for findings that demonstrate an ecological factor does moderate what was investigated in a current study. Lastly, researchers could refrain from speculating about the extent to which the generalizability of findings may be limited by ecological factors lacking empirical backing.

### **3. Arousal and Emotion Moderate Learning**

Emotion has become a "hot" (pun intended) topic on learning science. How might emotions relate to learning? I argue they are moderators not mediators.

Knowledge is fashioned by cognitively operating on information, e.g., generating an explanation for causal influences improves comprehension and memory (Bisra, Liu, Nesbit, Salimi, & Winne, 2018). Affect or emotion can moderate a process like self explanation and other cognitive operations applied to information in two ways. First, an affective state or emotional thought may be one factor a learner judges when deciding whether to apply a cognitive operation to particular information. This influences what a learner learns not because the learner is in a particular emotional state but because a cognitive process is or is not applied to particular information when the learner experiences a particular emotion.

Second, the contents of long-term memory are multidimensional. Learners can judge the accuracy, thoroughness, and reliability of their knowledge (although they sometimes err in the accuracy of their judgments). Knowledge in long-term memory is associated with a variety of other contents, for example, contextual descriptions about when a proposition was added to memory, what other knowledge it relates to and an affective stance regarding that proposition. Elaborations like these form a network, and characteristics of the cognitive network correlate to what learners perceive, learn and can recall. An affective experience or emotional thought is elaborative content. These experiences augment results of "cold" cognitive operations and are added into the network



of long-term memory. Cognitive operations are the causes of learning. Affect and emotional experiences are content that may moderate cognitive operations.

Several articles in this special issue set a stage for an intriguing deduction. A learner may not be aware a particular emotion is aroused while learning but, nonetheless, that state of arousal may influence cognition or its manifestation in interpersonal interactions. Thus, what learners learn may vary without the learner's awareness. What research has yet to explore is whether it is helpful to alert a learner that "now" is the time to regulate arousal. A complication to such a study is learners' capacities to change one affective state into another, that is, to regulate affect. Alerts about tacit emotions will be effective in proportion to this aptitude. Future research might probe how well learners can regulate affect and whether learners can learn to regulate affect to benefit learning.

#### **4. Puzzles about Proxy Measures of Emotion**

Physiological signals and facial displays are proxies for learners' experiences of emotions and other motivational constructs. Measures generated by instruments that represent these experiences are fundamentally different. Measures developed using physiological sensors used to detect arousal, for example, blood flow or EDA, rely on comparisons to a baseline. Researchers declare deviations from a subjectively set threshold to mark arousal or recovery. Deviations span time. In contrast, facial displays are measured by configurations of points on a face at a point in time. These measures are absolute; a configuration is matched or it is not. Time is irrelevant to the measurement. This contrast invites a question: What roles does time or rate of change play in a learner's perception of affect? Do learners perceive affect as variance in arousal over time or do they perceive affect as step function, on or off? What implications, if any, might this have for theorizing about affect and for learners approach to regulating affect?

##### **4.1 A Vexing Tautology**

Both kinds of proxies for affect are validated by asking learners to label their arousal. This creates a tautology. A researcher observes a measurement and asks the learner, "What is your emotion now?" Thereafter, that measurement is taken as a signal of the named emotion. How can the researcher be sure? The learner said so.

Why is this tautology an issue? A benefit of instrumentation used to gather proxy data is its unobtrusiveness. The learner does not have to be interrupted to report states of arousal or variation in affect. However, beyond instrumenting arousal, researchers are interested to identify valence. Today's state-of-the-art instruments can't do that without confronting the tautology just described.

#### **5. Is Reliability Relevant?**

When a learner reports and proxy measurements do not correlate highly, two cases need sorting. First, one or the other – the learner or the instrument – may be biased. Learners may be





reluctant to report some affects. Social demand does not disappear in the lab, work groups or classrooms. Also, instruments may be miscalibrated. Detecting and correcting bias in learners' reports requires ground truth. The previously mentioned tautology undermines faith there is ground truth.

A second case to consider when learners' reports do not highly correlate with proxy measurements is when the learner, the instrument or both are unreliable. Data generated by traditional instruments – responses to survey questions, answers to achievement test items, and the like – almost always prompt researchers to investigate the reliability of those data. Following Cronbach, Gleser, Randa and Rajaratnam (1972), reliability is a matter of identifying sources of variance in scores, identifying which of those factors (or facets) cannot be explained or controlled, and quantifying the contribution those “nuisance” sources make to variance in scores. That is, unreliability arises because nuisance factors introduce erratic variance in learners' reports or an instrument's measurements.

Some instruments generating proxy data about arousal, for example, sensors registering EDA, can be examined for reliability using other mechanical systems with known reliability. This helps to disassociate erratic variability in measurements researchers use to gauge learners' arousal. Any residual variance attributed to extraneous factors is then ascribed to the learner. But, if the learner is not biased (i.e., reliably misreports about arousal), can the learner be wrong in declaring their experience of affect? Again, the tautology occludes interpretation. Which source of data is the reliable source?

## **6. Final Thoughts: Applying Findings and the Evolution of Learning Science**

Learners' interpretations of states of arousal create their emotions and motivations. They will report emotions and motivations when asked. Researchers gather and interpret proxy data about these states. Both learners' reports and researchers' proxies correlate with what learners learn.

Learners' emotions and motivations do not create knowledge or change knowledge. Knowledge is created and amended when learners apply cognitive operations to information. Emotions and motivations are factors, mixed with other information, learners weigh in choosing which information to operate on and which cognitive operations to apply to selected information. According to this logic, emotions and motivations are moderators but not causes of learning. If this logic is correct, what are implications for helping learners become better learners?

I encourage research aim to identify conditions learners perceive that arouse them. Then, identify which of those conditions correlate positively and which correlated negatively with what learners learn. With a more precise map of relations between conditions learners perceive about their instructional context and learning results, instructional designs can more dependably be engineered to offer learners experiences in which learners can more productively self-regulate learning.

This account of learning leads to a potential surprise. Suppose conditions of instruction that arouse a particular learner are known. Suppose further conditions positively related to this learner's



achievements can be incorporated into the instructional context and those that negatively correlate with learning can be removed. Can learning science now ignore a learner's motivations and emotions. Is the ultimate goal of learning science finding dependable principles for engineering features of instruction that are devoid of constructs about inner states like emotions and motivations? Should learning science strive to locate theoretical constructs such as emotions and motivations to the category of epiphenomena?

This line of thinking explicitly and emphatically acknowledges students are individuals. Averaging out individual differences to describe effects at the level of a randomly formed group pervades research in learning science. I argue this is a mistake in attempts to model learners as self-regulating agents (Winne, 2017). Learners are agents (Winne, 2018). They choose how to learn. They are in control. What matters is their perception of instructional conditions. Each learner's perception emerges from a history of individual experiences about which conditions merit attention, what those conditions represent and predict about the present context, and what is a path to goals that context.

The surprise I reveal about learning science and guidance it can generate for designing instruction rests on the axiom that learners are agents. Instructional design therefore must be sensitive to, responsive to and supportive of learners as individuals. In this quest, instrumentation and methods like those described in this special issue are a boon. With energetic attention to methodological and interpretive issues affecting how data are interpreted, learning science carried out using methods and with instrumentation described in this special issue offers great opportunity to elevate understandings about each learner as an individual, about relationships between each learner's perceptions of their learning environment, and what learners learn in those contexts.

### **Key Points**

- Process data are better described in terms of shared subjectivity rather than as objective.
- Transformations of data change constructs and too often are not recognized in reporting effects.
- Concerns about ecological validity often lack empirical grounding.
- Emotion and motivations are moderators, not causes of learning.
- Grounding interpretations of process data on learners' reports is tautological, undermining traditional concerns about reliability.
- Surprisingly, progressive research relating process data to achievement may render emotions and motivations to the category of epiphenomena.

### **Acknowledgements**

Foundations for this article were developed with financial support provided over many years by the Social Sciences and Humanities Council of Canada and Simon Fraser University.





## References

- Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, *30*, 703-725.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed., pp. 13-103). New York: Macmillan
- Winne, P. H. (1982). Minimizing the black box problem to enhance the validity of theories about instructional effects. *Instructional Science*, *11*, 13-28
- Winne, P. H. (1983). Distortions of construct validity in multiple regression analysis. *Canadian Journal of Behavioural Science*, *15*, 187-202.
- Winne, P. H. (2017). Leveraging big data to help each learner upgrade learning and accelerate learning science. *Teachers College Record*, *119*(3), 1-24.
- Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In D. Schunk & J. Greene (Eds.), *Handbook of self-regulation of learning and performance*. (2<sup>nd</sup> ed., pp. 36-48). New York, NY: Routledge.
- Winne, P. H., Teng, K., Chang, D., Lin, M. P-C., Marzouk, Z., Nesbit, J. C., Patzak, A., Raković, M., Samadi, D., & Vytasek, J. (2019). nStudy: Software for learning analytics about processes for self-regulated learning. *Journal of Learning Analytics*, *6*, 95-106.