

# Chapter 2: A Practitioner’s Guide to Measurement in Learning Analytics - Decisions, Opportunities, and Challenges

Geraldine Gray,<sup>1</sup> Yoav Bergner<sup>2</sup>

<sup>1</sup> School of Informatics and Engineering, Technological University Dublin, Dublin Ireland

<sup>2</sup> Learning Analytics Research Network, New York University, New York City, USA

DOI: 10.18608/hla22.002

## ABSTRACT

What is our data measuring, why are we measuring it, and what can we infer from our measurements? These are key questions for models of learning, and the focus of this chapter. This chapter discusses the role of measurement in transitioning from predictive models of learning to models from which meaningful explanations about learning can be inferred. We consider how to associate latent constructs of learning with observable data from a variety of data sources relevant to learning contexts, illustrated with examples from recent LAK proceedings. We also review common sources of errors that arise with a variety of data collection instruments, and highlight the challenges and opportunities for progressing valid and reliable measurement of both learning itself and factors related to the learning process.

**Keywords:** Measurement, educational data sources, latent constructs, sources of error, explanatory models of learning

In the first edition of this handbook, the corresponding chapter on Measurement linked the foundational ideas of latent trait theory and methodology as they apply to learning analytics and educational data mining [5]. For this second edition, we have sought to supplement that work with more guidance for practitioners. We have thus structured the chapter in terms of decisions, opportunities, and challenges that practitioners face in using measurement methods for learning analytics. In particular, we look at measurement choices, and their consequences for inferring explanations from learning analytics models. The first section explores measurement more generally, and decisions related to why and what to measure. The second section looks at measurement choices for a selection of learning constructs, and the challenges and opportunities that arise from each choice.

## 1 DECIDING WHY AND WHAT TO MEASURE

### 1.1 Why measure? Understanding, explanation, optimization, and/or prediction

Practitioners often use the word “measure” synonymously with “observe”, including essentially all data collection. For the purpose of asking why and what we measure, with a lowercase *m*, there is no need yet for the kind of distinction that marks the statistical Measurement mod-

els of psychometricians, distinguished here with a capital *M*. Nevertheless, it is good practice to ask some *why* questions at an early stage in planning learning analytics projects. In particular, practitioners should be mindful of whether their ultimate goal is predictive or explanatory in nature. Findings that may serve predictive purposes well are not easily turned into explanatory results after the fact. Among the various definitions of learning analytics, most contain a purpose statement which references both “understanding” and “optimizing” (or “improving”) learning experiences. These words reinforce one another, and learning analysts pursue both goals. In practice, however, understanding and optimization do not always go hand in hand. We begin by clarifying some of these distinctions.

The function of understanding, which is used interchangeably with explanation, is necessarily bound up with theories of learning (and, more broadly, psychology, social-cognition, etc.) and even with value systems (i.e., the desirability of behaviors and other outcomes). Explanations of learning outcomes, unless very strictly behaviorist, inevitably appeal to concepts that are not directly observable (e.g., motivation, self-concept, aptitude). Understanding is usually labored and rarely simple. Explanation must admit challenges—alternative explanations—to the validity of its arguments. Optimizing or improving learning outcomes and environments need not be so. Optimization, however, must involve a step beyond prediction.

Frequently used in learning analytics research are various types of predictive modeling. (Authors use  $x$  and  $y$  to predict  $z$ .) (to cite chapter 3 in this edition [8]). Note that predictive analytics are not necessarily causal. As is often pointed out, one variable can be predictive of another when both have a common cause. For example, more time spent in a discussion forum of a course may be predictive of (that is, it may correlate positively with) more time spent using interactive simulations. But whatever lever might be used directly to get students to spend more time in the discussion forum is not necessarily going to increase simulation usage, or vice versa. We would most likely *explain* the correlation between these two observations by appealing to overall effort commitment and/or conscientiousness. In fact, if students truly have limited (but individually variable amounts of) time to allocate to a course, then forcing them to spend more time on one learning resource might, in principle, reduce the time they allocate elsewhere.

For prediction to be used for optimization on the student side, there must at least be a causal mechanism by which some design decision, adaptation, or intervention may be expected to change outcomes. It should be noted that causal relationships may still not rise to the level of explanations. Consider this: even a child knows that pressing on the rocker switch causes the ceiling lamp to light. But this is a far cry from understanding electric circuits or what to do if the light does not go on. Detached from a larger theoretical framework of mediators and moderators, causal findings in learning analytics may still guide future research. But optimization without explanation tends to be, at best, unsatisfying and, at worst, unethical. Computer algorithms that ignore the web of interconnected personal and social variables can perpetuate and exacerbate inequitable systems [53].

All of this is not to put down all predictive modeling. Indeed prediction or classification, as ends in themselves, often involves substantial and impressive technical progress. Optimization of the learning environment does not always require explanation. An illustrative example can be chosen from slightly outside the scope of learning analytics. Trained on large image data sets, computers today can identify dogs and fire hydrants with impeccable accuracy. Using deep neural network architectures, machines can even generate new, creative images of non-existent dogs. But does a computer with such capabilities “understand” the difference between a dog and a fire hydrant? Of course not. By contrast, a visually-impaired person understands that dogs are tail-wagging, domesticated wolves that develop strong bonds with humans who feed and care for them. But that won’t help in classifying a visual image that they can’t see clearly. One can know things that contribute to understanding while still struggling with specific tasks, and one can optimize performance in a specific task without a general understanding. It is tempting to point out deficiencies in the computer model—for example, it won’t be able to predict which one, the dog or the fire hydrant, is more likely to scratch itself or walk into the road. But of course a computer can be trained for those tasks too. Computer vision can be helpful for

everyone, visually impaired or not, and self-driving cars may in time prove safer than human drivers. Understanding, sense-making, and explanation, however, will remain distinctively human pursuits.

In the remainder of this chapter, we will become a bit stricter about what constitutes Measurement with a capital M. As we shall describe, Measurement is an emergent relationship between data and latent or hidden constructs that is mediated by a model. Insofar as “measures” are used in learning analytics for explanatory purposes, practitioners should be aware of several issues and challenges (sometimes called “validity threats”) that are pointed out in this chapter. We acknowledge that these issues may not apply uniformly to all data analyses, such as efforts to streamline or automate grading using machine learning methods.

## 1.2 What to measure? Learning constructs

The connection between data collected in a learning context and a construct of learning is not always direct. Learning analytics may be concerned, for example, with increases in student abilities or changes in student affect. Knowledge, ability, affect, and specific cases thereof are *learning constructs*. They are latent variables because they are not directly observable, so they must be inferred from directly observed indicators. It could even be said that learning constructs such as knowledge and ability are invented to explain patterns in observations, such as a tendency to solve problems correctly. Marks awarded for solving problems correctly in a test (test scores) or scale (survey) scores are directly observed indicators. Tests or surveys are instruments whose questions are considered to be Measurements of specific constructs. This is equivalent to saying that these constructs explain the observed data. However, there are limitations, some of which come down to common sense, about what should be considered a measure of what. For example, we might measure attendance and find it to be predictive of test scores. However, we do not consider attendance itself as a Measurement of ability. Given a Measurement model for the construct of conscientiousness, however, attendance might reasonably be considered a relevant indicator. Whether attendance is a high quality measure of conscientiousness, however, is still another matter.

Recent publications from Learning Analytics and Knowledge (LAK) conferences provide a sense of how the field uses Measurement. Some references are collected as elements in Table 1. Each paper is categorized by a column heading indicating a class of latent constructs (learning gains as well as traits, processes, and affective states, etc.) and a row heading representing the principle data sources for those constructs.

Determining if data collected in a learning context is a reasonable Measure of a construct of learning involves a number of steps. The first step is to identify the learning construct of interest. For example, a study may be generally interested in conscientiousness, or may be interested in a specific facet of conscientiousness like industriousness. The second step is to select an appropriate measure-

ment model for the construct, i.e. what can be measured (observed) as an indicator for the latent (unobservable) construct of interest? As illustrated by the examples cited in Table 1, there can be a number of measurement models to choose from. For example, *boredom* could be measured by a self-reported survey, third party observation, or by analysing images of facial expressions captured during the learning task. Each measurement model has its advantages, shortcomings, and sources of error, which will be explored later in the chapter.

The third step is implementing the measurement model as a data collection instrument. For example, what facets of facial expressions will be recorded to indicate *boredom*, and how frequently should features be sampled? The goal of measurement instruments is to capture a Measurement that is both **valid** and **reliable**. **Validity** refers to the interpretation of collected data as measures of the construct of interest. For example, do questionnaire answers or facial expression, actually measure *boredom*? For what intents and purposes? **Reliability** refers to the repeatability or consistency of the instrument observations. If validity is analogous to systematic error, then reliability is akin to random error. For example, how much range in facial feature detection might be attributed to the same level of boredom? Evaluating a Measurement instrument is often an iterative process of refinement and reevaluation. Some level of error is inevitable. For example, a systematic error could be caused by questionnaire items being interpreted differently in a particular context or culture that resulted in all responses underestimating *boredom*. Another source of error could be due to individuals' facial gestures varying in their level of expressiveness, resulting in random errors of both over- and under- estimates of *boredom*.

In sum, generating Measurement models of learning constructs involves a chain of methods for data collection, data cleaning, preprocessing, exploration and modelling. As the variety of chapters in this handbook testifies, there is a rich, eclectic mix of methods used in the field of learning analytics. The resulting methodology can be considered a chain of evidence from data to inference, as illustrated in Figure 1. Every step in the chain is a potential source of both error and alternative explanation. The next section explores some of these sources of error in more detail, specifically focusing on Measurements of constructs related to learning processes, learning gain, and potential data sources for each as exemplified by the elements in Table 1.

## 2 CHALLENGES AND OPPORTUNITIES

### 2.1 Measurement of Learning Process

As illustrated by the column headings in Table 1, a range of constructs are understood to influence the learning process. These include learner disposition, learner affect, pedagogical approach and epistemological beliefs [36, 27]. Measurements that capture aspects of the learning process are important in progressing explanatory models. The following paragraphs discuss a selection of measurement

models used to measure facets of the learning process, to highlight decisions and considerations relevant to their Measurement.

#### 2.1.1 Survey data, challenges and opportunities

Surveys are a data collection instrument for a variety of learning constructs. Using existing, validated survey instruments has the benefit of ensuring results can be compared and reproduced. In addition, tried and tested statistical techniques to assess internal validity (e.g. factor analysis) and internal reliability (e.g. Cronbach alpha or McDonalds Omega) are easily applied to survey items. A challenge with this measurement model is its inherent biases, particularly for self-reported scales. Sources of error include individuals or groups interpreting scale items differently, not remembering correctly, or individual perception being an under- or over- estimate of subjective measures such as abilities, emotions, or motivation levels [49]. In some cases, self-report measures are directly connected to the construct, such as when attitude surveys ask about the learner's enjoyment and perceived value of studying math. Other times, the target construct may be significantly moderated by the respondent's own perceptions, such as a survey that asks student's about their tendency to work well in a team.

#### 2.1.2 Trace data, challenges and opportunities

Trace data from educational technology has the advantage of removing the need for self-reporting, thus potentially eliminating biases inherent in survey data [49], as well as eliminating the effort in administering an additional data collection instrument. There is a wealth of data generated by educational technologies. Experimenting with a variety of static and dynamic features derived from trace data has generated relatively accurate predictive models in specific contexts. The challenge arises when attempting to draw inferences and explanations from these models. Recall the steps outlined earlier to evaluate observable data as a reliable indicator of a construct of learning, starting with defining the unobservable construct of interest. When analysis starts in the middle of these steps, with the observable data itself, working backwards to evaluate the measurement instrument as an indicator of a learning construct is problematic. This is because trace data reflects the instructional context that generated it. So the learning constructs it may measure, and the validity of that measurement, is dependent on how the technology was used in that instructional context. Reasonable validity and reliability in one context is unlikely to generalise to other contexts because working backwards from collected data to a measurement model is context specific. A good example of this from Table 1 is Motz et al. [36], who discuss the lack of portability of indices from VLE activity as a measure of behavioural engagement, based on an analysis of data from 829 courses. It's another side of the coin of "one model does not fit all" [17]. The evaluation of readily available trace data in one context does not fit all contexts.

For trace data to be considered a valid Measurement of a

Table 1: Citations sorted according to categories of constructs related to the learning process and 21st century skills

Data sources	Affect	Cognitive load	Collaborative learning	Non-cognitive traits	Behaviors	Domain knowledge
Image/Video	[46, 11]		[10, 52]		[10]	[25]
Text data	[9, 18]		[9, 32, 52]	[48]	[3, 15, 40, 39]	[4, 21, 26, 25]
Survey data	[20, 11]	[31]	[52]	[21, 45, 1]	[36]	
Trace data	[11]	[31]	[52]	[1, 20, 36]	[34, 36, 40, 45]	[26, 50]
Wearables/biometric	[11, 21]	[31, 46, 47]				
Network data			[41]			[43]

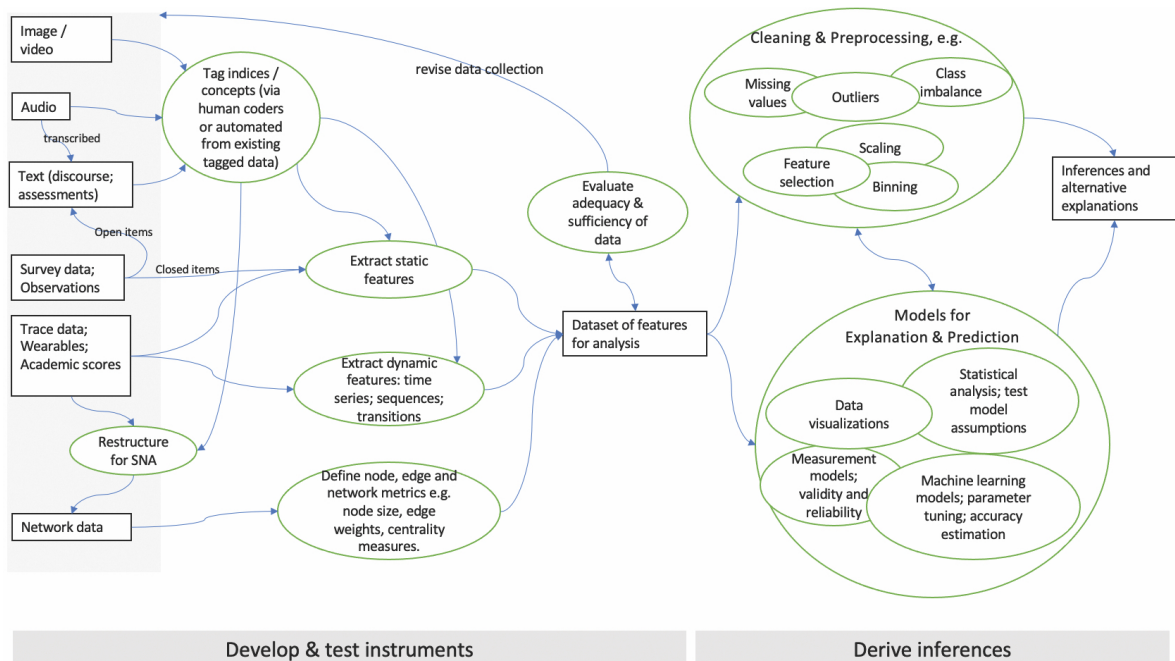


Figure 1: Chains of Evidence

learning construct, data collection should be preceded by identifying the learning constructs of interest, and defining the measurement model. For educational technology, this means deliberately designing the collection instrument (and so the consequential trace data it collects) around constructs of the learning process [24]. Examples from Table 1 include data from simulations using *Science Classroom Inquiry* that were designed around a specific pedagogical approach [40] so the instructional context is embedded in the tool. In another example, Harpstead et al. [20] configured the game *Decimal Point* to vary their construct of interest, *agency*. Simpler solutions, such as designing activities and resources on a VLE or MOOC to deliberately reflect a pedagogical approach are also viable (e.g. Matcha et al.[34]). In all these examples, the data collection instrument was configured to collect data about a latent construct of interest, increasing the likelihood of more generalisable estimates of construct validity from trace data.

While building instructional design into education technology can address model variance across pedagogical contexts, inferences should also consider variance due to learner contexts [44]. Trace data from electronic devices, such as wearables and image data, can capture data from contexts where learning is happening offline (e.g. face to face, or collaborative learning environments). They also collect data about the learner themselves. Therefore, such devices are a potentially useful addition to the landscape of trace data about learners and learning as discussed in (to cite Ochoa [38]). Biometric devices are measuring an observable construct directly (e.g. skin temperature). Image data requires some preprocessing, but libraries exist to automatically extract simple measurements from image data like posture, eye tracking and other motions. The challenge again arises when determining if the trace data is an indicator of an unobservable construct of learning. Validation typically uses manual coding and/or comparison with a second, validated data source, such as a validated questionnaire for the same construct. So one measurement model is validated with another, both of which have sources of error. Larmuseau et al. [31] provides an example of this. They found correlations between skin temperature and self-reported cognitive load in some instructional contexts only. So exploring the merit of such trace data as Measurements of constructs of the learning process offers opportunities for further research.

### 2.1.3 Text data, challenges and opportunities

Text data can capture the student voice directly, with the potential to provide different, and potentially richer insights than both surveys and trace data, as discussed in chapters 5, 10 and 11 or this text [2, 19, 12]. Indicators from text data can relate to the learning process and learning gain. So how does text data map to measurement? Models of learning require input data to be structured. Therefore, unstructured text data must be converted to structured data where features are the constructs of interest, and the data are based on counts of those features. Counts can be simple, such as term or phrase counts. More interestingly

for explanatory modelling, counts of features derived from language usage that evidence learning constructs can also be extracted from text. Tools like *Linguistic Inquiry and Word Count* (LIWC), and *Coh-Metrix*, automatically extract linguistic measures such as psychological processes (e.g. affect) and aspects of writing cohesion respectively. Natural language is inherently imprecise and its meaning can be subjective. In spite of this, a number of studies have confirmed the validity of automatically extracted Measures from these tools when the assessment/writing brief reflects the constructs of interest, for example, Kovanović et al. [29] and Jung & Wise [23]. Where an automated feature extraction tool is not available for a construct of interest, training a model to extract more complex features from text requires a training dataset of text that has been manually coded (labelled). For example, Stone et al. [48] trained a model to infer a selection of non-cognitive traits from a 150-word essay about extracurricular activity, and reported good agreement with human coders of the same essays. Although Eagan et al. [14] warns of the potential for high Type I errors when using human coders to assess reliability in learning contexts.

### 2.1.4 Temporal considerations

Regardless of the measurement model, many constructs of learning have a temporal aspect. For example, cycles between positive and negative emotions can have a positive impact on the learning process compared to maintaining a consistent emotion [16]. Similarly, a change in student behaviour over time might be more insightful than a snapshot or aggregate of their behaviour. So, as well as verifying indicators from a measurement instrument, an additional step in the evidence chain may be warranted to define, measure, and model transitions between states of a construct.

## 2.2 Measurement of Learning Gain

Learning gain may refer to growth in knowledge, skills, or competencies during a period of interest. For specific content domains, such as algebra, developing reliable measures is a straightforward if laborious process. However, as the learning domain becomes more complex, so do the Measurement challenges [54, 27]. Assessment of competencies such as ways of thinking and ways of working, is a challenge facing educators more generally [30, 35]. Indeed, the difficulty in settling on agreed terminology related to non-cognitive dimensions (defining the constructs) evidences the range of opportunities that exist in this under explored space [22]. As with *learning process*, technology offers interesting opportunities for Measurement of non-cognitive skills (see, e.g., [42, 13]).

Another consideration when measuring learning gain is the period during which the learning was gained. Ideally, an instrument would measure learning gained as a change over time [51], for example, differences in pre- and post-test scores as discussed in [37]. Learning analytics models more frequently use existing post-test scores or assessment aggregates such as end of term grade (without

a pre-test). While these scores reflect measurement in a real context, there is an assumption that the learning was gained during the period of analysis.

The granularity of measurement also impacts on the resulting interactions captured by a model. Proficiency in coarse grained or complex learning outcomes is a continuous variable. Reporting learning gained as alpha grades aims to compensate for errors inherent in the subjective nature of marking assessments. While this is good practice from pedagogical perspective (see, e.g. Kohn [28]), from a data modelling perspective, this reduces the granularity of the information content to an ordinal scale with somewhat arbitrary bin boundaries. Data preparation for modelling academic performance may reduce granularity further by dichotomising to a label such as pass/fail. There is information loss when a continuous attribute is discretized. For example, resulting analysis underestimates linear relationships between the original, continuous variables and other independent variables of interest, thus increasing the chance of type II errors [33, 7]. Dichotomisation may also introduce main effects not present in the original, continuous variables [33].

### 3 CONCLUSION

This chapter has considered a variety of sources of observable data that offer potential indicators of unobservable constructs of learning, and discussed some of the challenges of using observable data to measure latent constructs. As was said in the introduction, explanations of models of learning must acknowledge these challenges and sources of error, and consider the resulting implications on explanations that are inferred from models of the data.

Sources of error do not end with the measurement model. Every method applied to the data during cleaning, pre-processing, operationalization choices, feature selection, modelling, parameter tuning and estimates of model fit can add additional sources of error [6]. The resulting model will inevitably include bias as models are based on the data that is available, which is incomplete. There will be subgroups of learners missing from the data. For the learners that are included, there will be mediators, moderators and confounders not captured that explain some of the model variance. Some gaps in the data may be obvious to us and so easy to identify. Other gaps could be related to factors that impact on learning, or categories of students, we haven't thought to consider yet.

So do we give up on Measurement? No, we accept the sources of error as part of a robust argument evaluating all methods used, to ensure measurement, methodology and resulting models and inferences are honestly critiqued. The key point is that we know that our models aren't perfect, and we interpret the data in full knowledge of its limitations. Overtime, as the body of robust evidence builds around Measurement of learning and resulting optimisations and explanations, we can progress as a field.

### REFERENCES

- [1] Stephen J. Aguilar and Clare Baek. "Motivated information seeking and graph comprehension among college students". In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. Tempe, Arizona: ACM, 2019, pp. 280–289.
- [2] Laura K. Allen, Sarah C. Creer, and Püren Öncel. "Natural Language Processing as a Tool for Learning Analytics - Towards a Multi-Dimensional View of the Learning Process". In: *The Handbook of Learning Analytics*. Ed. by Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron. 2nd ed. Vancouver, Canada: SoLAR, 2022. Chap. 5, pp. 46–53. ISBN: 978-0-9952408-3-4. URL: <https://www.solaresearch.org/publications/hla-22/hla22-chapter5/>.
- [3] Laura K. Allen, Caitlin Mills, Cecile Perret, and Danielle S. McNamara. "Are you talking to me? Multi-dimensional language analysis of explanations during reading". In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 116–120.
- [4] David Azcona, Piyush Arora, I-Han Hsiao, and Alan Smeaton. "user2code2vec: Embeddings for profiling students based on distributional representations of source code". In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 86–95.
- [5] Yoav Bergner. "Measurement and its Uses in Learning Analytics". In: *Handbook of Learning Analytics*. Ed. by Charles Lang, George Siemens, Alyssa Wise, and Dragan Gasevic. First. Society for Learning Analytics Research (SoLAR), May 2017, pp. 35–48. ISBN: 978-0-9952408-0-3. URL: <https://solaresearch.org/hla-17/hla17-chapter3>.
- [6] Yoav Bergner, Geraldine Gray, and Charles Lang. "What does methodology mean for learning analytics?" In: *Journal of Learning Analytics* 5.2 (Aug. 5, 2018), pp. 1–8–1–8. ISSN: 1929-7750. DOI: 10.18608/jla.2018.52.1. URL: <https://learning-analytics.info/index.php/JLA/article/view/6164>.
- [7] Yoav Bergner, Charles Lang, and Geraldine Gray. "A focus on methodology in learning analytics: Building a structurally sound bridge discipline". In: *Workshop on MLA/BLAC at the 7th International Conference on Learning Analytics & Knowledge*. 2017.
- [8] Christopher Brooks and Craig Thompson. "Predictive Modelling in Teaching and Learning". In: *The Handbook of Learning Analytics*. Ed. by Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron. 2nd ed. Section: 3. Vancouver, Canada: SoLAR, 2022, pp. 29–37. ISBN: 978-0-9952408-3-4. URL: <https://www.solaresearch.org/publications/hla-22/hla22-chapter3/>.

- [9] Heeryung Choi, Nia Dowell, Christopher Brooks, and Stephanie Teasley. "Social comparison in MOOCs: Perceived SES, opinion, and message formality". In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 160–169.
- [10] Mutlu Cukurova, Qi Zhou, Daniel Spikol, and Lorenzo Landolfi. "Modelling collaborative problem-solving competence with transparent learning analytics: is video data enough?" In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK '20: 10th International Conference on Learning Analytics and Knowledge. Frankfurt Germany: ACM, Mar. 23, 2020, pp. 270–275. ISBN: 978-1-4503-7712-6. DOI: 10 . 1145 / 3375462 . 3375484. URL: <https://dl.acm.org/doi/10.1145/3375462.3375484>.
- [11] Sidney D'Mello, Ed Dieterle, and Angela Duckworth. "Advanced, analytic, automated (AAA) measurement of engagement during learning". In: *Educational psychologist* 52.2 (2017), pp. 104–123.
- [12] Nia Dowell and Vitomir Kovanović. "Modeling Educational Discourse with Natural Language Processing". In: *The Handbook of Learning Analytics*. Ed. by Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron. 2nd ed. Vancouver, Canada: SoLAR, 2022. Chap. 11, pp. 105–119. ISBN: 978-0-9952408-3-4. URL: <https://www.solaresearch.org/publications/hla-22/hla22-chapter11/>.
- [13] Nia MM Dowell, Yiwen Lin, Andrew Godfrey, and Christopher Brooks. "Exploring the relationship between emergent sociocognitive roles, collaborative problem-solving skills, and outcomes: A group communication analysis." In: *Journal of Learning Analytics* 7.1 (2020), pp. 38–57.
- [14] Brendan Eagan, Jais Brohinsky, Jingyi Wang, and David Williamson Shaffer. "Testing the reliability of inter-rater reliability". In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 2020, pp. 454–461.
- [15] Elaine Farrow, Johanna Moore, and Dragan Gašević. "Analysing discussion forum data: a replication study avoiding data contamination". In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 170–179.
- [16] Klaus Fiedler and Susanne Beier. "Affect and cognitive processes in educational contexts". In: *International handbook of emotions in education*. Educational psychology handbook series. New York, NY, US: Routledge/Taylor & Francis Group, 2014, pp. 36–55. ISBN: 978-0-415-89501-9.
- [17] Dragan Gasevic, Shane Dawson, Tim Rogers, and Danijela Gasevic. "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success". In: *The Internet and Higher Education* 28 (Jan. 2016), pp. 68–84. ISSN: 10967516. DOI: 10 . 1016 / j . iheduc . 2015 . 10 . 002. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1096751615300038>.
- [18] Shay A. Geller, Nicholas Hoernle, Kobi Gal, Avi Segal, Amy X. Zhang, David Karger, Marc T. Facciotti, and Michele Igo. "Confused and beyond: detecting confusion in course forums using students' hashtags". In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 2020, pp. 589–594.
- [19] Andrew Gibson and Antonette Shibani. "Natural Language Processing - Writing Analytics". In: *The Handbook of Learning Analytics*. Ed. by Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron. 2nd ed. Vancouver, Canada: SoLAR, 2022. Chap. 10, pp. 96–104. ISBN: 978-0-9952408-3-4. URL: <https://www.solaresearch.org/publications/hla-22/hla22-chapter10/>.
- [20] Erik Harpstead, J. Elizabeth Richey, Huy Nguyen, and Bruce M. McLaren. "Exploring the subtleties of agency and indirect control in digital learning games". In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 121–129.
- [21] Xiao Hu, Fanjie Li, and Runzhi Kong. "Can background music facilitate learning? Preliminary results on reading comprehension". In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 101–105.
- [22] Srecko Joksimovic, George Siemens, Yuan Elle Wang, M. O. Z. San Pedro, and Jason Way. "Editorial: Beyond cognitive ability". In: *Journal of Learning Analytics* 7.1 (Apr. 3, 2020), pp. 1–4–1–4. ISSN: 1929-7750. DOI: 10 . 18608 / jla . 2020 . 71 . 1. URL: <https://learning-analytics.info/index.php/JLA/article/view/7076>.
- [23] Yeonji Jung and Alyssa Friend Wise. "How and how well do students reflect? multi-dimensional automated reflection assessment in health professions education". In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 2020, pp. 595–604.
- [24] Kirsty Kitto, John Whitmer, Aaron E Silvers, and Michael Webb. "Creating Data for learning analytics ecosystems". In: (2020). Publisher: Society for Learning Analytics Research (SoLAR). URL: <https://www.solaresearch.org/core/creating-data-for-learning-analytics-ecosystems/>.
- [25] Beata Beigman Klebanov, Anastassia Loukina, John Lockwood, Van Rynald T. Licalde, John Sabatini, Nitin Madnani, Binod Gyawali, Zuwei Wang, and Jennifer Lentini. "Detecting learning in noisy data: the case of oral reading fluency". In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK '20: 10th International Confer-

- ence on Learning Analytics and Knowledge. Frankfurt Germany: ACM, Mar. 23, 2020, pp. 490–495. ISBN: 978-1-4503-7712-6. DOI: 10.1145/3375462.3375490. URL: <https://dl.acm.org/doi/10.1145/3375462.3375490>.
- [26] Beata Beigman Klebanov, Anastassia Loukina, Nitin Madnani, John Sabatini, and Jennifer Lentini. “Would you?: Could you? On a tablet? Analytics of children’s eBook reading”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. LAK19: The 9th International Learning Analytics & Knowledge Conference. Tempe AZ USA: ACM, Mar. 4, 2019, pp. 106–110. ISBN: 978-1-4503-6256-6. DOI: 10.1145/3303772.3303833. URL: <https://dl.acm.org/doi/10.1145/3303772.3303833>.
- [27] Simon Knight, Simon Buckingham Shum, and Karen Littleton. “Epistemology, pedagogy, assessment and learning analytics”. In: *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. LAK ’13. New York, NY, USA: Association for Computing Machinery, Apr. 8, 2013, pp. 75–84. ISBN: 978-1-4503-1785-6. DOI: 10.1145/2460296.2460312. URL: <https://doi.org/10.1145/2460296.2460312>.
- [28] Alfie Kohn. *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A’s, Praise, and Other Bribes*. Houghton Mifflin Harcourt, 1999.
- [29] Vitomir Kovanović, Srećko Joksimović, Negin Mirriahi, Ellen Blaine, Dragan Gašević, George Siemens, and Shane Dawson. “Understand students’ self-reflections through learning analytics”. In: *Proceedings of the Eight International Conference on Learning Analytics & Knowledge (LAK’18)*. LAK ’18. Sydney, NSW, Australia: ACM, 2018. DOI: 10.1145/3170358.3170374.
- [30] Patrick C. Kyllonen. “Measurement of 21st century skills within the common core state standards”. In: *Invitational Research Symposium on Technology Enhanced Assessments*. 2012, pp. 7–8.
- [31] Charlotte Larmuseau, Pieter Vanneste, Piet Desmet, and Fien Depaepe. “Multichannel data for understanding cognitive affordances during complex problem solving”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 61–70.
- [32] Yiwen Lin, Nia Dowell, Andrew Godfrey, Heeryung Choi, and Christopher Brooks. “Modeling gender dynamics in intra- and interpersonal interactions during online collaborative learning”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 431–435.
- [33] Robert C. MacCallum, Shaobo Zhang, Kristopher J. Preacher, and Derek D. Rucker. “On the practice of dichotomization of quantitative variables.” In: *Psychological methods* 7.1 (2002), p. 19.
- [34] Wannisa Matcha, Dragan Gašević, Nora’Ayu Ahmad Uzir, Jelena Jovanović, and Abelardo Pardo. “Analytics of learning strategies: Associations with academic performance and feedback”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 461–470.
- [35] Sandra Milligan. “Standards for developing assessments of learning using process data”. In: *Reimagining University Assessment in a Digital World*. Springer, 2020, pp. 179–192.
- [36] Benjamin Motz, Joshua Quick, Noah Schroeder, Jordon Zook, and Matthew Gunkel. “The validity and utility of activity logs as a measure of student engagement”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. LAK19. New York, NY, USA: Association for Computing Machinery, Mar. 4, 2019, pp. 300–309. ISBN: 978-1-4503-6256-6. DOI: 10.1145/3303772.3303789. URL: <https://doi.org/10.1145/3303772.3303789>.
- [37] Benjamin A. Motz, Paulo F. Carvalho, Joshua R. de Leeuw, and Robert L. Goldstone. “Embedding experiments: Staking causal inference in authentic educational Contexts”. In: *Journal of Learning Analytics* 5.2 (Aug. 5, 2018), pp. 47–59–47–59. ISSN: 1929-7750. DOI: 10.18608/jla.2018.52.4. URL: <https://learning-analytics.info/index.php/JLA/article/view/5802>.
- [38] Xavier Ochoa. “Multimodal Learning Analytics - Rationale, Process, Examples, and Direction”. In: *The Handbook of Learning Analytics*. Ed. by Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron. 2nd ed. Vancouver, Canada: SoLAR, 2022. Chap. 6, pp. 54–65. ISBN: 978-0-9952408-3-4. URL: <https://www.solaresearch.org/publications/hla-22/hla22-chapter6/>.
- [39] Melanie Peffer, David Quigley, Liza Brusman, Jennifer Avena, and Jennifer Knight. “Trace data from student solutions to genetics problems reveals variance in the processes related to different course outcomes”. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 2020, pp. 47–52.
- [40] Melanie E. Peffer and Niloofar Ramezani. “Assessing epistemological beliefs of experts and novices via practices in authentic science inquiry”. In: *International Journal of STEM Education* 6.1 (Jan. 22, 2019), p. 3. ISSN: 2196-7822. DOI: 10.1186/s40594-018-0157-9. URL: <https://doi.org/10.1186/s40594-018-0157-9>.
- [41] Oleksandra Poquet, Liubov Tupikina, and Marc Santolini. “Are forum networks social networks? A methodological perspective”. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 2020, pp. 366–375.



- [42] Tenelle Porter, Diego Catalán Molina, Lisa Blackwell, Sylvia Roberts, Abigail Quirk, Angela Lee Duckworth, and Kali Trzesniewski. “Measuring mastery behaviors at scale: The persistence, effort, resilience and challenge-seeking task (PERC)”. In: *Journal of Learning Analytics* 7.1 (2020), pp. 5–18.
- [43] Chen Qiao and Xiao Hu. “Measuring knowledge gaps in student responses by mining networked representations of texts”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 275–279.
- [44] Joshua Quick, Benjamin Motz, Jamie Israel, and Jason Kaetzel. “What college students say, and what they do: aligning self-regulated learning theory with behavioral logs”. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK '20. New York, NY, USA: Association for Computing Machinery, Mar. 23, 2020, pp. 534–543. ISBN: 978-1-4503-7712-6. DOI: 10.1145/3375462.3375516. URL: <https://doi.org/10.1145/3375462.3375516>.
- [45] Joseph M. Reilly and Chris Dede. “Differences in student trajectories via filtered time series analysis in an immersive virtual world”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 130–134.
- [46] Kshitij Sharma, Zacharoula Papamitsiou, Jennifer Olsen, and Michail Giannakos. *Predicting learners' effortful behaviour in adaptive assessment using multi-modal data*. Mar. 8, 2020.
- [47] Namrata Srivastava, Sadia Nawaz, Jason M. Lodge, Eduardo Velloso, Sarah Erfani, and James Bailey. “Exploring the usage of thermal imaging for understanding video lecture designs and students' experiences”. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 2020, pp. 250–259.
- [48] Cathlyn Stone, Abigail Quirk, Margo Gardener, Stephen Hutt, Angela L. Duckworth, and Sidney K. D'Mello. “Language as thought: Using natural language processing to model noncognitive traits that predict college success”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 320–329.
- [49] Dirk Tempelaar, Bart Rienties, and Quan Nguyen. “Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application”. In: *PLOS ONE* 15.6 (June 12, 2020). Ed. by Vitomir Kovanovic, e0233977. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0233977. URL: <https://dx.plos.org/10.1371/journal.pone.0233977>.
- [50] Khushboo Thaker, Paulo Carvalho, and Kenneth Koedinger. “Comprehension factor analysis: Modeling student's reading behaviour: Accounting for reading practice in predicting students' learning in MOOCs”. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019, pp. 111–115.
- [51] Jan D. Vermunt, Sonia Ilie, and Anna Vignoles. “Building the foundations for measuring learning gain in higher education: a conceptual framework and measurement instrument”. In: *Higher Education Pedagogies* 3.1 (Jan. 2018), pp. 266–301. ISSN: 2375-2696. DOI: 10.1080/23752696.2018.1484672. URL: <https://www.tandfonline.com/doi/full/10.1080/23752696.2018.1484672>.
- [52] Hana Vrzakova, Mary Jean Amon, Angela Stewart, Nicholas D. Duran, and Sidney K. D'Mello. “Focused or stuck together: multimodal patterns reveal triads' performance in collaborative problem solving”. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK '20. New York, NY, USA: Association for Computing Machinery, Mar. 23, 2020, pp. 295–304. ISBN: 978-1-4503-7712-6. DOI: 10.1145/3375462.3375467. URL: <https://doi.org/10.1145/3375462.3375467>.
- [53] Betsy Anne Williams, Catherine F. Brooks, and Yotam Shmargad. “How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications”. In: *Journal of Information Policy* 8 (2018), p. 78. ISSN: 23815892. DOI: 10.5325/jinfopoli.8.2018.0078. URL: <https://www.jstor.org/stable/10.5325/jinfopoli.8.2018.0078>.
- [54] Mark Wilson, Isaac Bejar, Kathleen Scalise, Jonathan Templin, Dylan Wiliam, and David Torres Irribarra. “Perspectives on methodological issues”. In: *Assessment and teaching of 21st century skills*. Springer, 2012, pp. 67–141.