

# What Does Methodology Mean for Learning Analytics?

Yoav Bergner<sup>1</sup>, Geraldine Gray<sup>2</sup>, Charles Lang<sup>3</sup>

## ABSTRACT

The diversity of substantive expertise and analytical techniques spanned by learning analytics can be a challenge for understanding and interrogating methodological choices. But active pursuit of methodological clarity is essential to productive multivocality. This was the focus of the LAK '17 Methodology in Learning Analytics workshop from which the idea for this special section arose. Discussions were framed using a building metaphor of developing structurally sound bridges between disciplinary methods. Strong arguments are buttressed by supporting evidence and warrants but also engage with potential alternative explanations. Papers in this special section further the discussion and encompass a range of relevant approaches including sensitivity analysis, model evaluation, model fit, causal inference, and visualization as a methodology. All advance the case of structural integrity in this bridge discipline and highlight the importance of adequate descriptions of methodological considerations to aid our common understanding.

### Keywords

Methodology, epistemology, evaluation, sensitivity analysis, bridge discipline

**Published:** 05.08.18

Corresponding author: <sup>1</sup>Email: [yoav.bergner@nyu.edu](mailto:yoav.bergner@nyu.edu) Address: NYU-LEARN, Learning Analytics Research Network, New York University, 82 Washington Square East, Pless Hall 7th Floor, New York, NY 10003, USA ORCID ID: 0000-0001-7738-4290

<sup>2</sup>Email: [geraldine.gray@itb.ie](mailto:geraldine.gray@itb.ie) Address: Department of Informatics, Institute of Technology Blanchardstown, Blanchardstown Road North, Dublin, D15 YV78, Ireland ORCID: 0000-0002-9694-2558

<sup>3</sup>Email: [charles.lang@tc.columbia.edu](mailto:charles.lang@tc.columbia.edu) Address: Box 118, 525 West 120th Street, New York, NY 10027 USA

## 1. Framing the Role of Methodology in Learning Analytics

“You keep using that word. I do not think it means what you think it means.” — Inigo Montoya

The idea for this special section emerged in the first Workshop on Methodology in Learning Analytics at LAK 2017 (Bergner, Lang, & Gray, 2017a; 2017b). That gathering addressed two related challenges common to new interdisciplinary fields, namely boundary finding and epistemological identity. Drawing the domain boundary of learning analytics is a broad work in progress beyond the scope of a specialized workshop. As a narrower substitute, the question “What does methodology mean for learning analytics?” seemed more tractable. We noted that Clow (2013) had described learning analytics’ “eclectic approach” to picking up methods from various disciplines as one that “facilitates rapid development and the ability to build on established practice and findings, but it — to date — lacks a coherent, articulated epistemology of its own” (p. 686). Convening the workshop, then, was a way of recasting the search for a coherent epistemology in terms of coherent approaches to methodology. The workshop was also an effort to address Suthers and Verbert’s (2013) notion of productive multivocality. Researchers from diverse fields may see fundamental incompatibilities from behind their respective disciplinary boundaries. Or they can try to span the diversity of views to build new knowledge. We have pursued methodology as a necessary tool in paving the way forward.

The words “methods” and “methodology” are sometimes used interchangeably. However, as we use the terms here, methods are specific instances or classes of analytic approaches such as regression, classification, ethnography, network analysis, and design research. Methodology refers to the detailed study of these methods, including their assumptions, properties, sensitivities, and justifications for use. All substantive researchers must engage in some amount of methodology, but some researchers will choose to linger more on its formal questions. Participants in our workshops, which have now run for two years, acknowledge the importance of a pluralistic approach when it comes to methods. Learning analytics, they feel, should not draw sharp boundaries around “our accepted methods” (at least not yet). However, there has also been a reluctance

to accept the idea that anything goes. The common sense that emerged was, “do your thing, but make it defensible.” Defensibility, as applied to arguments and justifications, is often at the heart of methodology. But defensibility invokes a metaphor that may not be the best choice for conceptualizing the role of methodology in the learning analytics community.

Linguists Lakoff and Johnson (1980) drew attention to the argument-as-war metaphor that is deeply embedded in the English language. Arguments can be attacked and defended, they can be weak or strong. They have winners and losers. Focusing on our relatively young and multivocal community of practice (Suthers & Verbert, 2013), we have attempted elsewhere (Bergner, Lang, & Gray, 2017a) to highlight an alternative metaphor, also described by Lakoff and Johnson (1980), of argument as building. Arguments can fall apart or be buttressed. The building metaphor seemed particularly appropriate for a young interdisciplinary field such as learning analytics. We pointed to the danger of building a bridge discipline such as ours on structurally weak arguments. We proposed that methodologists may be thought of as quality assurance engineers in the bridge-building enterprise of learning analytics, stress testing the roadways, trusses, and wire ropes that connect educational technologists, psychologists, data scientists, learning scientists, substantive experts in various educational domains, and measurement specialists. The arguments that scholars in learning analytics make need to be understood well enough by peer reviewers to probe their weaknesses. Identification of weaknesses helps, of course, to avoid spurious inferences. But this task, which often falls on the finger-wagging methodologist, also helps researchers by pushing them to clarify and strengthen their arguments.

The quality-assurance work of methodology interrogations tends to follow a particular pattern. To characterize it, we find it helpful to refer to Toulmin’s (2003) argument structure, which extends naturally to the structure of academic papers. A representation of the Toulmin model is shown in Figure 1. The end of an argument is a claim, which is supported by data/observations. The observations alone, however, do not typically suffice for non-trivial arguments. There must be a warrant for justifying the claim from the observations, and the warrant itself is backed by additional evidence. Finally, alternative explanations and rebuttal evidence serve to qualify a claim. (Conversely, as suggested above, pointing out the weakness of alternative explanations can be a way of strengthening the original warrant.) The model is simple but remarkably general.

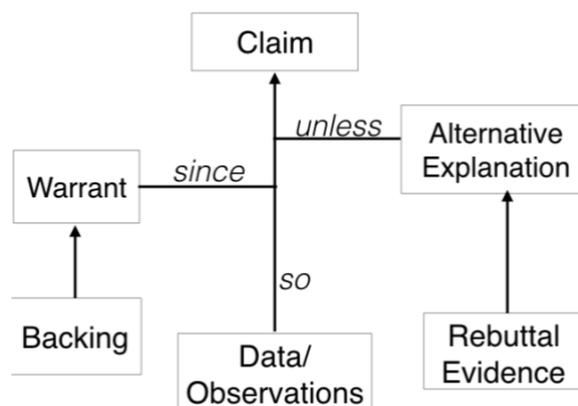
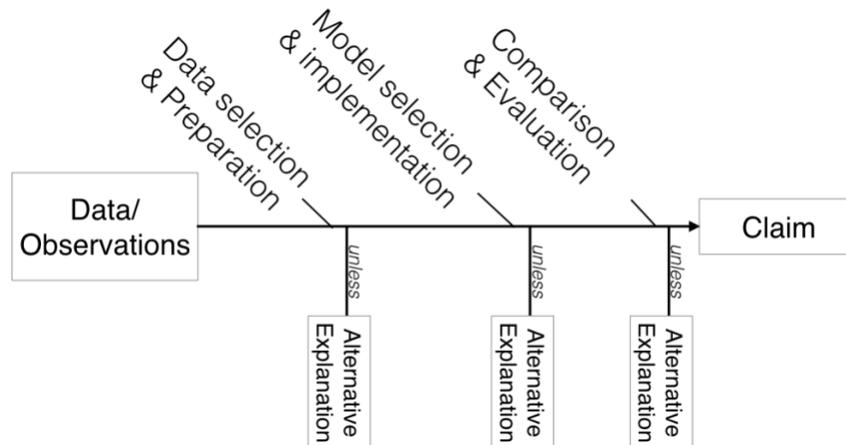


Figure 1: Toulmin model of a reasoning argument.

Of course, a paper is usually composed of a chain of arguments rather than a single one. A simplified representation of an archetypal learning analytics paper might involve the chain illustrated in Figure 2. It is not implied that a paper is written in this order, but rather that the logical path from data/observations to a claim (about learners, technologies, pedagogies, etc.) will typically involve warrants justifying data selection and preparation (possibly multiple such steps), model selection and implementation, and evaluation. For each of these links in the chain of argument, alternative explanations exist that potentially undermine the claim. For example, data processing may be criticized as being overly sensitive to operationalization choices (see, for example, Fincham, Gašević, & Pardo [2018] and Gardner & Brooks [2018], both in this issue). Paper authors who aim to build a strong argument are likely to engage with these alternative explanations. Methodologists, in particular though, make it their business to understand the branch points in this diagram.



**Figure 2:** Sample chain of arguments in a learning analytics paper. Methodology is concerned with the warrants and, in particular, the alternative explanations for claims.

In sum, the substantive researcher wishes to make a convincing argument by appealing to supporting evidence and, ideally, by dismissing some potential alternative explanations. The methodologist plays the role of interrogator, challenging the warrants and propping up the alternative explanations. Both are essential in advancing the science of learning. Whereas substantive research questions establish “what” it is that we want to know (direct path to claim in Figure 1), methodology resides at the interaction between *the what* and *the how* (perpendicular branch points for *since* and *unless*). We call it an interaction, as distinguished from simply an additional consideration, because the answer to *the how* question can depend rather finely on the answer to *the what* question. For example, methodologists may happily dwell on the question of whether an outcome measure of interest as conceived of by the researcher is continuous or discrete, because a mismatch between analysis methods and variable type can lead to an increase in the rate of Type I and Type II errors (MacCallum, Zhang, Preacher, & Rucker, 2002). Applying a classifier to a continuous variable after a median split is a common example of problematic practice. The methodologist asks: “What would have happened if the research design, model selection, or evaluation had been carried out differently? Would the same conclusions have been drawn?” We suggest that this type of interrogation is particularly important and helpful in learning analytics given its multivocality and the breadth of techniques that practitioners use.

## 2. Specific Issues for Methodology in Learning Analytics: Papers in This Section

With respect to statistical and machine-learning models, learning analytics is for the most part an applied art/science. Conference and journal papers in this field rarely develop brand new algorithms, instruments, or statistical estimators, but they use them, often in novel contexts. For example, applications of classifier algorithms to student affect (e.g., Bosch & Paquette [2018], in this issue) or network analysis to discussion forum activity (Fincham et al. [2018], in this issue) raise issues that were not always anticipated and accounted for in the foundational work on classifier and network methods. Other applied fields have gone through their own hand-wringing over methodology. Of particular note is the econometrics work of Kennedy (2002), in which the author provides ten rules/commandments for practitioners.<sup>1</sup> Kennedy’s rules carry over remarkably well to the field of learning analytics, which is perhaps still too young for a distinct set of its own. He defends a nuanced position on data mining as exploratory analysis (Commandment #7) that anticipates well the contemporary perspective. Another example will likely seem relevant to learning analytics practitioners: “Thou shalt confess in the presence of sensitivity” (Commandment #10). Sensitivity refers to the effect that researcher’s methodological decisions may have on reported outcomes.

Learning analytics research has spawned a number of recent papers emphasizing the sensitivity of quantitative analyses to data collection and variable operationalization choices. Examples have included the effects of selection bias (Brooks, Chavez, Tritz, & Teasley, 2015), results of time-on-task analyses (Kovanović et al., 2015), operationalization of discussion forum usage (Bergner, Kerr, & Pritchard, 2015), and evaluation of student models (Pelánek, Rihák, & Papoušek, 2016). Fincham, Gašević, and Pardo’s paper in this issue — “From Social Ties to Network Processes: Do Tie Definitions Matter?” — is another exemplar of sensitivity analysis studies. The authors stress-test one of the bridges of learning analytics, connecting the field to that of

<sup>1</sup> See also references within for prior art on the ten commandments of statistical inference.

statistical network analysis. This paper is a timely contribution, as the application of network methods to data from course discussion forums is becoming widespread.

Methodological choices are particularly important in network analyses, because discussion forum data are not “naturally” represented as network structures. Consider, as a contrasting example, relations in social media networks such as Facebook and Twitter. A symmetric friendship relation exists between Leonard and Janice if they have identified each other as friends on Facebook. A directed tie from Leonard to Janice exists if Leonard follows Janice on Twitter. The ties in each of the two network definitions above have a natural one-to-one relationship with the affordances of the original platforms. However, discussion forums are different. Students who participate by posting and replying in a thread do not necessarily declare to whom they are responding. Thus, some operationalization choices need to be made to define the network given the data, and, as Fincham et al. point out, previous studies have used different definitions. The methodological question here becomes, do substantive inferences about the network and about relationships between person-in-network properties and other person behaviours (e.g., academic performance) depend on the operationalization choices? Spoiler alert: yes they do! By pointing out the magnitude of these consequences, this paper helps to highlight a responsibility for learning analytics researchers who use network methods.

In “Learning as it Happens: A Decade of Analyzing and Shaping a Large-Scale Online Learning System,” Brinkhuis, Savi, Hofman, Coomans, van der Maas, and Maris (2018) adapt some established protocols from educational measurement (for example, in large-scale testing) to the monitoring of *Math Garden*, an adaptive learning environment. This should be a welcome bridge for learning analytics. In high-stakes testing, great care is taken to validate the uses of test scores for particular purposes, such as ranking or credentialing. Poor test design can have serious consequences (Bergner, 2017). Learning systems like *Math Garden*, where the stakes are relatively low, can operate under less stringent standards. However, there are still negative consequences of sub-optimal design. Learning gains might actually be diminished in comparison to non-adaptive alternatives. There are different ways to evaluate the design of an adaptive system; for example, one may use experimental designs (see, for example, Motz, Carvalho, de Leeuw, & Goldstone [2018], in this issue). Alternatively, when a measurement model exists, then a type of internal verification arises from model fit analysis. Model misfit can flag problematic design, and it can sometimes help identify unanticipated learner behaviours.

In the early days of Khan Academy, for example, mastery was defined by a fixed number of correct consecutive practice attempts. This heuristic was not based on a measurement model, and thus model fit analysis would not have been possible. (The current mastery scale is a bit more complex). In other cases, such as in cognitive tutors (Ritter, Anderson, Koedinger, & Corbett, 2007), quantitative models of performance are indeed validated against observed behaviour during practice. If there is a mismatch between expected and observed outcomes, the system itself may be adjusted (Martin, Mitrovic, Koedinger, & Mathan, 2011). Either practice problems may be regrouped into different skill groupings — a structural change — to the knowledge map, or model parameters may be updated.

Model fit is a methodology — distinct from the method of adaptive model application — for assessing whether a measurement model is performing as expected. In terms of the language of claims, warrants, and alternative explanations, the role of model fit analysis in adaptive learning environments might be described as follows: An adaptation rule is a prescription typically used along with an implied claim that it is better than a fixed sequence of practice. Moreover, the reason the adaptive algorithm is supposed to be better is that it makes correct inferences from behaviours to learner understanding/mastery, which is not directly observed. But (alternative explanation) what if the inferences are wrong, and the adaptive content selection is essentially arbitrary, or possibly even worse than chance? In this issue, Brinkhuis et al. (2018) elaborate on several important model fit methodologies in the context of *Math Garden*, including the following: 1) visual inspection of residual differences between predicted and observed scores; 2) mixture modelling of residual differences to obtain corrected estimates of root mean square error; and 3) contingency table analysis, with constraints based on model specification.

“Embedding Experiments: Staking Causal Inference in Authentic Educational Contexts” by Motz, Carvalho, de Leeuw, and Goldstone (2018) also addresses evaluation of system design and interventions, arguing for greater use of embedded experimental design to support causal inferences. Learning in context is influenced by many factors (Richardson, Abraham, & Bond, 2012), making identification of specific causal inferences challenging. Controlling for factors via experimentation in a laboratory setting changes this context and so affects the dependent variable of interest. Conducting learning experiments in context is not new (Imai, King, & Stuart, 2008), but is not common in learning analytics publications, as argued by Motz et al. Their paper addresses a range of issues that may have precluded researchers from considering an embedded (in context) experimental design in a learning context, and it discusses a variety of approaches for conducting such embedded experiments.

A chain of evidence from data/observations to a claim is not sufficient to infer a causal relationship between variables in the data, even with appropriate warrants and due regard paid to arguments against alternative explanations. Many models of academic performance report similar, good predictive accuracy based on a variety of variables (e.g., class tests, online activity, demographic data), without explaining causes of strong or weak academic performance (Tempelaar, Rienties, & Nguyen,

2017). In addition, evidence from educational psychology lacks consistency on which mediators explain relationships between observable variables and academic performance; results vary depending on which variables were included in a study (Gray, McGuinness, Owende, & McCarthy, 2014). Motz et al. offer a bridge between research and practice that supports causal inference, and bring us a step closer to Siemens and Long's (2011) claim that learning analytics could "penetrate the fog of uncertainty on how to allocate resources" in the planning of learning activities. It is debatable how much of that fog has been cleared to date, particularly with respect to actionable intelligence.

Whether or not learning analytics might take us some way toward clearing the fog of uncertainty is an open question, but it can safely be assumed that without foregrounding methodological choices in learning analytics we run the risk of generating more doubt. Without the ability to differentiate inferential conclusions based on how they were constructed, results will multiply without any visible organizational super-structure. A paper by Alhadad (2018) in this issue approaches the problem of foregrounding methodological choices with respect to data visualization, using principles of cognitive psychology to frame the consequences of different visual choices for both making inferences and communicating ideas. This approach has relevance for generating bridges between many of the disciplines that relate to learning analytics as all utilize visualization in some form. It also provides a specific bridge between cognitive science and other disciplines by drawing on the accumulated wisdom of that field to inform methodological choices. In Alhadad's article, cognitive psychology is the grounding for how learning analytics can best leverage visualization for communicating results and as a methodological tool. In establishing this framework, Alhadad suggests two uses of alternative explanations. First, alternative ways of depicting data can communicate ideas less or more effectively. Second, visualization becomes a research methodology in itself when alternative visualizations can lead to different epistemological conclusions. By addressing the limits of human psychology with respect to both attention and cognition, learning analytics might be able to best position itself as a maximally useful field for education by both communicating ideas effectively and gaining novel insights.

Returning to Kennedy's Commandment #7 ("Understand the costs and benefits of data mining"), it has been argued that data mining methods are something of a fishing exercise to see what patterns a particular data set can offer (Hand, 1998). Evaluation of resulting models is done with the same sample that generated the model (even if using a different subset of the sample). Model fit statistics, comparing observed with expected results, do not apply. Instead, performance metrics compare predicted values with actual (observed) values in a labelled test dataset that is subject to the same systematic and unsystematic sources of error as the model's training data. A discussion of alternative explanations is needed to address these potential sources of error.

Accepting data mining methods as exploratory, there are a range of commonly used performance metrics that differ in how they quantify the level of agreement between predicted values and actual values. These metrics thus provide different perspectives on estimates of model performance. Pelánek's (2015) review of AUC and RMSE (for continuous and binary outcomes) highlighted not only the importance of discussing the rationale for choice of metric, but also of providing details of their computation, yet another source of variation in metric results. Another article in this issue — "Metrics for discrete student models: Chance levels, comparisons, and use cases" by Bosch and Paquette (2018) — extends this conversation with an evaluation of metrics for models making discrete predictions of discrete labels (specifically, F1, precision, recall, and Cohen's Kappa). In datasets of discrete student outcomes, class imbalance is a common problem, resulting in potentially high performance for chance level models. The authors distinguish between imbalance in the dataset and imbalance in the predicted classes when evaluating performance metrics. The paper highlights the importance of interpreting model accuracy with respect to the chance level for each metric reported. It also questions the limitations of relying on a single evaluation metric, as is typically deployed in wrapper methods for feature selection, for example.

Gardner and Brooks (2018; also in this issue) interrogate the application of performance metrics as well but with respect to the pitfalls of model comparison. Like Bosch and Paquette, their paper explores the bridge between learning analytics and machine learning. Gardner and Brooks express concern that the state of predictive modelling in learning analytics is somewhat of a wild west, without rigorous reporting, for example, on model construction, hyper-parameter tuning, and adjustments for multiple comparisons in significance tests.

Model comparison is in itself a test of alternative explanations (alternative models). However, Gardner and Brooks go further in their methodological argument. They anticipate the following hypothetical claims:

After collecting vast amounts of clickstream data from a MOOC, a wide range of behaviour features were extracted and a variety of predictive modelling architectures were used to predict final course grades. The best model had a root mean squared error of X. Moreover, the most explanatory features of that model were Y. Therefore, it is inferred that further research be conducted to establish whether Y behaviours causally mediate performance in learners.

The alternative explanation: Many of the models were statistically indistinguishable from each other. Though they may differ in feature weights, there is no basis for understanding the importance of these features. The authors emphasize that standardizing criteria for judging the utility of models is a crucial stress test of the whole enterprise of predictive modelling. After reviewing a number of statistical approaches, Gardner and Brooks propose a Bayesian approach that distinguishes between practical equivalence and significant differences in model average performance. They apply these methods for ranking the performance of 96 models derived from a MOOC dataset.

### 3. Discussion and Considerations for Future Practice

While research results published in other domains provide some justification for methodological choices in learning analytics, there are specific characteristics inherent to learning analytics that merit attention, as highlighted by the papers in this issue. Learning is a complex latent trait, influenced by contexts that cannot be replicated in a lab setting, as discussed by Motz et al. and Fincham et al. Time-honoured methods for accrediting learning based on assessments and examination are controversial (Knight, Buckingham Shum, & Littleton, 2013), but the alternatives are noisy indicators of academic competencies. This has implications for interpretation of student models and their evaluation metrics (discussed by Brinkhuis et al., Bosch & Paquette, and Gardner & Brooks) and for inferring reliable, actionable intelligence from model outputs (discussed by Motz et al.). In addition, student behaviours and their knowledge, skills, and competencies are fluid, so data footprints of learning activities are limited temporally; Brinkhuis et al. note some of the challenges of real-time learning analytics. Although not unique to an educational context, experimenting with interventions raises ethical concerns, as addressed in Motz et al., who argue for more inclusion of experimental design in learning analytics and propose that technology can enable a more scientific interrogation of causal influences. Brinkhuis et al. provide an example of how this can be done in a large-scale learning system. Both papers recognize the value of large data collections from a live learning environment that can capture complexities inherent in the learning process, while acknowledging the potential biases that can be introduced by data collection. The use of learning analytics in practice requires appropriate representation of results for stakeholders who are not methodological experts, as discussed by Alhadad in considerations for data visualization choices. Together, these six papers highlight the importance of differentiating between accurate reporting of facts and understanding that learning is more complex than what facts and models alone can capture. George Box's (1979) assertion that "all models are wrong, but some are useful" is particularly relevant to learning analytics.

Where do we go from here? It is important that a robust conversation around methodology be developed within the learning analytics community. Just as any number of methodological weak links can undermine the structural integrity of an argument, so they can undermine a whole bridge discipline. Learning analytics is particularly susceptible given the breadth of techniques practitioners use. As our field borrows and combines methods liberally from other domains, it is challenging for readers and reviewers to be fluent with all of them. It is therefore important that authors communicate their methodological choices with transparency and appropriate citations, so that a trail of accepted methodology can be built up. This both facilitates the utilization of methods and the ability for other researchers to critique and improve on those methods. Ground rules for conducting and communicating analyses should be debated rigorously and will take time to coalesce. Papers such as those found in this special section, as well as events such as the methodological workshops at LAK, provide forums for public discussion of methodology within learning analytics and for thinking through complex problems in a systematic fashion.

One of the great strengths of learning analytics is its multivocality. While the current special section has a preponderance of quantitative methods, the selection of these papers should not be seen as an indication that the guest editors believe that other methodological avenues are less worthy of inclusion. Balancing the diversity of methodologies and having them communicate with and about each other in productive ways is a great challenge faced by methodologists in our growing field. It behooves all of us involved in the enterprise to contribute positively to this conversation and not retreat back inside our disciplinary ramparts. By concentrating on methodology as a means of building strong bridges, learning analytics may realize the promise that has brought so many diverse fields together.

### Declaration of Conflicting Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) declared no financial support for the research, authorship, and/or publication of this article.

## References

- Alhadad, S. S. J. (2018). Visualizing data in learning analytics: Cognitive research in support of inference-making. *Journal of Learning Analytics*, 5(2), 60–85. <http://dx.doi.org/10.18608/jla.2018.52.5>
- Bergner, Y. (2017). Measurement and its uses in learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 34–48). SoLAR, Society for Learning Analytics Research. <http://dx.doi.org/10.18608/hla17.003>
- Bergner, Y., Kerr, D., & Pritchard, D. E. (2015). Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes. In O. C. Santos et al. (Eds.), *Proceedings of the 8<sup>th</sup> International Conference on Educational Data Mining (EDM2015)*, 26–29 June 2015, Madrid, Spain (pp. 234–241). International Educational Data Mining Society.
- Bergner, Y., Lang, C., & Gray, G. (2017a). Workshop on methodology in learning analytics. *Proceedings of the 7<sup>th</sup> International Conference on Learning Analytics and Knowledge (LAK '17)*, 13–17 March 2017, Vancouver, BC, Canada (pp. 500–501). New York: ACM. <http://dx.doi.org/10.1145/3027385.3029427>
- Bergner, Y., Lang, C., & Gray, G. (2017b). A focus on methodology in learning analytics: Building a structurally sound bridge discipline. In Y. Bergner et al. (Eds.), *Joint Proceedings of the Workshop on Methodology in Learning Analytics (MLA) and the Workshop on Building the Learning Analytics Curriculum (BLAC)*, co-located with 7<sup>th</sup> International Learning Analytics and Knowledge Conference (LAK 2017), 13–14 March 2017, Vancouver, BC, Canada, *CEUR Workshop Proceedings*, Vol. 1915, paper #1. <http://ceur-ws.org/Vol-1915/paper1.pdf>
- Bosch, N., & Paquette, L. (2018). Metrics for discrete student models: Chance levels, comparisons, and use cases. *Journal of Learning Analytics*, 5(2), 86–104. <http://dx.doi.org/10.18608/jla.2018.52.6>
- Box, G. E. (1979). Robustness in the strategy of scientific model building. In Launer & Wilkinson (Eds.), *Robustness in statistics* (pp. 201–236). Cambridge, MA: Academic Press.
- Brinkhuis, M. J. S., Savi, A. O., Hofman, A. D., Coomans, F., van der Maas, H. L. J., & Maris, G. (2018). Learning as it happens: A decade of analyzing and shaping a large-scale online learning system. *Journal of Learning Analytics*, 5(2), 29–46 <http://dx.doi.org/10.18608/jla.2018.52.3>
- Brooks, C., Chavez, O., Tritz, J., & Teasley, S. (2015). Reducing selection bias in quasi-experimental educational studies. *Proceedings of the 5<sup>th</sup> International Conference on Learning Analytics and Knowledge (LAK '15)*, 16–20 March 2015, Poughkeepsie, NY, USA (pp. 295–299). New York: ACM. <http://dx.doi.org/10.1145/2723576.2723614>
- Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, 18(6), 683–695.
- Fincham, E., Gašević, D., & Pardo, A. (2018). From social ties to network processes: Do tie definitions matter? *Journal of Learning Analytics*, 5(2), 9–28 <http://dx.doi.org/10.18608/jla.2018.52.2>
- Gardner, J., & Brooks, C. (2018). Evaluating predictive models of student success: Closing the methodological gap. *Journal of Learning Analytics*, 5(2), 105–125. <http://dx.doi.org/10.18608/jla.2018.52.7>
- Gray, G., McGuinness, C., Owende, P., & McCarthy, A. (2014). A review of psychometric data analysis and applications in modelling of academic achievement in tertiary education. *Journal of Learning Analytics*, 1(1), 75–106. <http://dx.doi.org/10.18608/jla.2014.11.5>
- Hand, D. J. (1998). Data mining: Statistics and more? *The American Statistician*, 52(2), 112–118.
- Imai, K., King, G., & Stuart, E. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2), 481–502. <http://dx.doi.org/10.1111/j.1467-985X.2007.00527.x>
- Kennedy, P. E. (2002). Sinning in the basement: What are the rules? The ten commandments of applied econometrics. *Journal of Economic Surveys*, 16(4), 569–589. <http://dx.doi.org/10.1111/1467-6419.00179>
- Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Epistemology, pedagogy, assessment and learning analytics. *Proceedings of the 3<sup>rd</sup> International Conference on Learning Analytics and Knowledge (LAK '13)*, 8–12 April 2013, Leuven, Belgium (pp. 75–84). New York: ACM. <http://dx.doi.org/10.1145/2460296.2460312>
- Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S. J. D., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. *Proceedings of the 5<sup>th</sup> International Conference on Learning Analytics and Knowledge (LAK '15)*, 16–20 March 2015, Poughkeepsie, NY, USA (pp. 184–193). New York: ACM. <http://dx.doi.org/10.1145/2723576.2723623>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40. <http://dx.doi.org/10.1037/1082-989X.7.1.19>
- Martin, B., Mitrovic, A., Koedinger, K. R., & Mathan, S. (2011). Evaluating and improving adaptive educational systems with learning curves. *User Modeling and User-Adapted Interaction*, 21(3), 249–283. <http://dx.doi.org/10.1007/s11257-010-9084-2>
- Motz, B. A., Carvalho, P. F., de Leeuw, J. R., & Goldstone, R. L. (2018). Embedding experiments: Staking causal inference

- in authentic educational contexts. *Journal of Learning Analytics*, 5(2), 47–59 <http://dx.doi.org/10.18608/jla.2018.52.4>
- Pelánek, R. (2015). Metrics for evaluation of student models. *Journal of Educational Data Mining*, 7(2), 1–19.
- Pelánek, R., Rihák, J., & Papoušek, J. (2016). Impact of data collection on interpretation and evaluation of student models. *Proceedings of the 6<sup>th</sup> International Conference on Learning Analytics and Knowledge (LAK '16)*, 25–29 April 2016, Edinburgh, UK (pp. 40–47). New York: ACM. <http://dx.doi.org/10.1145/2883851.2883868>
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387. <http://dx.doi.org/10.1037/a0026838>
- Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), 249–255. <http://dx.doi.org/10.3758/BF03194060>
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30.
- Suthers, D., & Verbert, K. (2013). Learning analytics as a middle space. *Proceedings of the 3<sup>rd</sup> International Conference on Learning Analytics and Knowledge (LAK '13)*, 8–12 April 2013, Leuven, Belgium (pp. 2–5). New York: ACM. <http://dx.doi.org/10.1145/2460296.2460298>
- Tempelaar, D. T., Rienties, B., & Nguyen, Q. (2017). Towards actionable learning analytics using dispositions. *IEEE Transactions on Learning Technologies*, 10(1), 6–16. <http://dx.doi.org/10.1109/TLT.2017.2662679>
- Toulmin, S. E. (2003). *The uses of argument, Updated edition*. Cambridge, UK: Cambridge University Press.