Is Learning Data in the Right Shape?

Anthony E. Kelly
US National Science Foundation
George Mason University
akelly1@gmu.edu

ABSTRACT: In this short thought-piece, I attempt to capture the type of freewheeling discussions I had with our late colleague, Mika Seppälä, a research mathematician from Helsinki. Mika, not being a psychometrician or learning scientist, was blissfully free from the design constraints that experts sometimes ingest, unwittingly. I also draw on delightful conversations with the German research mathematician, Heinz-Otto Peitgen, a polyglot whose work includes advances in medical imaging and explorations in fractal geometry for K–12 students. Together, they taught me to reconsider foundational assumptions about learning, how to describe it, and how to grow it. Accordingly, I use this set of papers as a prompt for examining assumptions that numerical precision ensures scientific insight, that linear models best capture growth in learning, and that relaxing a fixation with time (exemplified by the reification of pre- and post-testing) might open up new topologies for describing, predicting, and promoting learning in its myriad manifestations.

Keywords: Learning, modelling, linearity, complexity theory

1   TOOLS FOR REPRESENTING DATA

Which mathematical tools are powerful for analyzing data on learning? For many education and social science researchers, typical quantitative tools include natural numbers, lines of best fit for scatterplots of coordinate points, and (comparisons of) measures of central tendency.

1.1   Numbers as Points on a Line

Many researchers routinely assume that "numbers" faithfully represent social and learning phenomena and that these numbers represent interval (or ratio) scales. However, distinctions drawn between nominal, ordinal, interval, and ratio data are often lost.

For example, in what sense is a "score" on a test a number? Accepting that all items on a 20-item test are not cognitively or semantically interchangeable, there are the 184,756 different ways that a score of "10 out of 20" can be generated. Thus, in what sense does a score of "10" represent a unique knowledge state for a learner? In what sense is a score of "10" diagnostic (i.e., to which pertinent set of the 184,756 options does it refer)? What can be inferred by the clustering of students who each scored "10" on the test? Further, in what sense is it valid to compare two groups who each scored an average of "10" and to argue that no differences exist between the groups? This problem is compounded when a

test requires sophisticated reasoning (e.g., Semak, Dietz, Pearson, & Willis, 2017), and when scores from disparate tests are combined to generate a course grade.

When we add scores on a test or generate means, we assert that a linear relationship is the appropriate geometric expression for modelling learning. However, is it the case that a student who scored 66 (on a 100-item test) knows "three times more" than the person who scored 22? Is the five-point difference between the scores of 55 and 60 equal, phenomenologically, to the difference between scores of 15 and 20, or 90 and 95?

It is beyond the scope of this paper to examine the use of numbers to describe behaviors in greater detail, but the interested reader may wish to explore the work of Tatsuoka classifying learners using cognitive task analyses employing rule space (Tatsuoka, 2009), the use of partially ordered set theory (Tatsuoka & Ferguson, 2003), and related methods that describe knowledge spaces (e.g., Heller, Stefanutti, Anselmi, & Robusto, 2015).

1.2   Associations

While curve fitting, time series, and trend analyses have been available for many decades, we often rely on straight lines to capture the shape of education data. However, assumptions of linearity may unconsciously blind our perceptions and predetermine our conclusions. For example, a low or zero correlation may suggest "no relationship" between variables. Yet, when curvilinear data are present, a simple Pearson correlation will incorrectly represent the phenomenon (see the first figure at https://en.wikipedia.org/wiki/Correlation_and_dependence). For a compelling example where disparate datasets have the exact same non-zero correlation coefficient, see Anscombe (1973).

2   MIKA SEPPÄLÄ AND THE SHAPE OF DATA

Discussions such as these with our late colleague, Mika Seppälä, led to three NSF awards (i.e., NSF Award Numbers: 1252625, 1338509, and 1450501). The most recent of these awards squarely approached the generative topic *the shape of educational data*. This grant supported a meeting in Fairfax, Virginia, from which this set of papers emanated.

As a research mathematician, Mika encouraged us to adopt tools other than points and lines. He favoured Riemann surfaces, and recommended that we examine the work of topologists such as Buser at Lausanne, Carlsson at Stanford, and Harer at Duke. In this vein, we find near-neighbour ideas proposed by fellow mathematicians Buser and Semmler (2017) and Munch (2017).

In this short piece, I extend the playful conversations that began with Mika and speculate on how the "shape of educational data" might illuminate some of the papers' points. None of these speculations should be considered a criticism of any paper; rather, I hope that they spur generative conversations. Indeed, the treatment by Caprotti of Markov graphs (2017) suggests that the following exploration may not be too fanciful.

3   THE TURN FROM POINTS TO SPACES

Ostrow, Wang, and Heffernan (2017) reported that an analog approach to partial credit provided more insight on learning than a binary scoring approach (i.e., correct/incorrect scoring only). This finding suggests that assessment is better viewed as sampling from a relevant knowledge space rather than a collection of binary switches that privileges point estimates. Minstrell showed that correct/incorrect scoring may punish learning along an entire learning trajectory (e.g., DeBarger, Ayala, Minstrell, Kraus, & Stanford, 2009). To further explore the knowledge space for assessment, the interested reader is directed to the work of Messick (1994) on validity, Lesh and colleagues on model-eliciting activities (Diefes-Dux, Hjalmarson, Miller, & Lesh, 2008), Mislevy's (2009) work on evidence-based design of assessments, and Schaffer's work on epistemic network analysis (Shaffer, Collier, & Ruis, 2016).

3.1   Playing with "Ribbons," Orbits, Attractors, and Phase Spaces

Buser and Semmler (2017) describe students' different educational tracks as tracing trajectories through a set of bifurcating cylinders. These cylinders are similar to subway paths marking the beginning to the end of a journey, explicitly bound to the variable of time.

In this vein, imagine that the primary shape that describes a domain expert's view of the content of a course is represented by a ribbon. Is this metaphor, if the content is judged uniformly difficult, the ribbon lies flat. If the course introduces difficult content at first (e.g., to "weed out" students), less challenging content toward the middle, and increasingly challenging material toward the end, the ribbon would trace a rising inclined plane, followed by a plateau, ending as a rising inclined plane. A number of other possible surfaces (e.g., staircases) may occur to the reader. Indeed, since the content of courses is complex, a set of ribbons may be required. For example, Pauna (2017) lists nine online assessments of calculus competencies that we may imagine reflect the content of the course (from factual recall to information transfer; see p. 13). Thus, for each student there may be a unique ribbon, tracing different pathways with different gradients through the course material (compare Pauna, 2017, on student pathways).

3.2   Assessment-of-Progress Ribbons

We can see from Pauna (2017) and from Caprotti (2017) that a student can take many pathways through the course resources: traversing quizzes, workshops, lectures, and other materials. For students, the actual course difficulty will be an interaction between the content and a range of individual and social factors (e.g., prior instructional history, readiness to learn, socioeconomic factors) (e.g., Gašević, Dawson, Rogers, & Gašević, 2016).

Thus, before a course begins, and once it is underway, we may predict the shape of the course trajectories for different ability students (e.g., via Bayesian updates based on their prior instructional histories, and covariates). Thus, an online course that was easily traversed by most students would have surface gradients consistent with a flat plain with an attractor of a passing grade. However, for a set of weaker students, their prior and emerging behaviors might predict the rapid emergence of a "basin" in the topology of the course predicting drop out or failure.

4   SOCIAL LEARNING, TUTORING, AND TRAVERSING

The contours of ribbons for failing students could change for each student in response to support from networks of students or from dedicated mentors (e.g., Treisman, 1992). We learn from Pauna (2017) and Caprotti (2017) that a comprehensive approach to modelling learning analytics should be responsive to individual, dyad–, group– and student–instructor interventions (see also Ayoubi, Pezzoni, & Visentin, 2017).

From Wang and Kelly (2017), we learn that video frames can be time-stamped and meta-tagged to be searchable by students and researchers. Further, videos can be organized in content-sensitive clips, and annotated by peers, teaching assistants or faculty. And, video segments can be interspersed with quizzes or other assessments (e.g., using the quizzes from Gage, 2017).

Thus, with strategic interventions by tutors or mentors, and by judicious use of course-support materials, the changing topology of a course may positively diverge from the emerging predictions (i.e., the "basin" may resolve itself for weaker students).

5   INTERSECTING CONTENT AND ASSESSMENT SPACES

We can now return to the techniques that describe knowledge spaces and ask anew if content in the instructional materials and assessment domains describe mutually intersecting surfaces. Ideally, course content ontologies, assessment material constructs, and student readiness indicators should mutually interpenetrate to advance student learning (see Gašević, Jovanović, Pardo, & Dawson, 2017). For example, if formative assessments in the calculus course measured only factual recall or the videos for certain topics were missing, basins predicting failure would appear in any shared content/assessment surface.

For example, let's focus on the mental rotation measure and its low correlation with final grades in the paper by Hart, Daucourt, and Ganley (2017). The authors wrote, "We also found it surprising that mental rotation was not an important predictor (or even a strong correlate) of final grade in Calculus II" (p. 146). However, the relationship between spatial abilities and STEM learning is complex, and different mathematical sub-constructs might relate to spatial abilities, but not be captured by a final grade (e.g., Stieff & Uttal, 2015; Uttal et al., 2013). Since Uttal and colleagues also argue that spatial abilities are malleable, targeted interventions related to spatial reasoning (justified by a task analysis of the course materials) might increase the correlation between spatial abilities and learning. In other words, the course and assessment design may not be sophisticated enough to adequately analyze and support the expression of students' abilities.

6   MOVING FORWARD

In addition to the suggestions above for reconsidering the shape of educational data prompted by our colleague Mika Seppälä, the reader is encouraged to attend conferences on learning analytics (e.g., those supported by SOLAR), to track investments related to the recent NSF 10 Big Ideas (especially the

157

ones on harnessing data and the human technology frontier[1]), and to review sources such as Foster, Ghani, Jarmin, Kreuter, and Lane (2017).

ACKNOWLEDGEMENTS

REFERENCES

Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician, 27*(1), 17–21. http://dx.doi.org/10.1080/00031305.1973.10478966

Ayoubi, C., Pezzoni, M., & Visentin, F. (2017). At the origins of learning: Absorbing knowledge flows from within the team. *Journal of Economic Behavior & Organization, 134*, 374–387. http://dx.doi.org/10.1016/j.jebo.2016.12.020

Buser, P., & Semmler, K.-D. (2017). Study paths, Riemann surfaces, and Strebel differentials. *Journal of Learning Analytics, 4*(2), 62–77. http://dx.doi.org/10.18608/jla.2017.42.7

Caprotti, O. (2017). Shapes of educational data in an online calculus course. *Journal of Learning Analytics, 4*(2), 78–92. http://dx.doi.org/10.18608/jla.2017.42.8

DeBarger, A. H., Ayala, C., Minstrell, J. A., Kraus, P., & Stanford, T. (2009). *Facet-based progressions of student understanding in chemistry* (Chemistry Facets Technical Report 1)*.* Menlo Park, CA: SRI International.

Diefes-Dux, H., Hjalmarson, M. A., Miller, T., & Lesh, R. (2008). Model-eliciting activities for engineering education. In J. S. Zawojewski, H. Diefes-Dux, & K. Bowman (Eds.), *Models and modeling in engineering education: Designing experiences for all students* (pp. 17–36). Rotterdam, Netherlands: Sense Publishers.

Foster, I., Ghani, R., Jarmin., R. S., Kreuter, F., & Lane, J. (Eds.). (2017). *Big data and social science*. New York: CRC Press/Taylor & Francis.

Gage, M. E. (2017). Methods of interoperability: Moodle and WeBWorK. *Journal of Learning Analytics, 4*(2), 22–35. http://dx.doi.org/10.18608/jla.2017.42.4

Gašević, D., Dawson, S., Rogers, T., & Gašević, D., (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. http://dx.doi.org/10.1016/j.iheduc.2015.10.002

Gašević, D., Jovanović, J., Pardo, A., Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics, 4*(2), 115–130. http://dx.doi.org/10.18608/jla.2017.42.10

Hart, S. A., Daucourt, M., & Ganley, C. M. (2017). Individual differences related to college students' course performance in calculus II. *Journal of Learning Analytics, 4*(2), 131–155.

---

[1] https://www.nsf.gov/about/congress/reports/nsf_big_ideas.pdf

158

http://dx.doi.org/10.18608/jla.2017.42.11

Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, *80*(4), 995–1019. http://dx.doi.org/10.1007/s11336-015-9457-x

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments**.** *Educational Researcher,* *23***(**2), 13–23. http://dx.doi.org/10.3102/0013189X023002013

Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. In R. L. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 83–108). Charlotte, NC: Information Age Publishing.

Munch, E. (2017). A user's guide to topological data analysis. *Journal of Learning Analytics, 4*(2), 47–61. http://dx.doi.org/10.18608/jla.2017.42.6

Ostrow, K. S., Wang, Y., & Heffernan, N. T. (2017). How flexible is your data? A comparative analysis of scoring methodologies across learning platforms in the context of group differentiation. *Journal of Learning Analytics, 4*(2), 93–114. http://dx.doi.og/10.18608/jla.2017.42.9

Pauna, M. (2017). Calculus courses' assessment data. *Journal of Learning Analytics, 4*(2), 12–21. http://dx.doi.org/10.18608/jla.2017.42.3

Semak, M. R., Dietz, R. D., Pearson, R. H., & Willis, C. W. (2017). Examining evolving performance on the Force Concept Inventory using factor analysis. *Physics Review Physics Education Research*, *13,* http://dx.doi.org/10.1103/PhysRevPhysEducRes.13.019903.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data**.** *Journal of Learning Analytics***,** *3*(3), 9–45. http://dx.doi.org/10.18608/jla.2016.33.3

Stieff, M., & Uttal, D. (2015). How much can spatial training improve STEM achievement? *Educational Psychology Review, 27*(4), 607–615. http://dx.doi.org/10.1007/s10648-015-9304-8

Tatsuoka, C., & Ferguson, T. (2003). Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), 65*(1), 143–157. Retrieved from http://www.jstor.org/stable/3088831

Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. New York: Routledge.

Treisman, U. (1992). Studying students studying calculus: A look at the lives of minority mathematics students in college. *College Mathematics Journal 23*(5), 362–372.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. http://dx.doi.org/10.1037/a0028446

Wang, S. P., & Kelly, W. (2017). Video-based big data analytics in cyberlearning. *Journal of Learning Analytics, 4*(2), 36–46. http://dx.doi.org/10.18608/jla.2017.42.5