Workshop on Educational Data Visualization

Nirmal Patel
Playpower Labs
nirmal@playpowerlabs.com

Derek Lomas
Delft University of Technology
j.d.lomas@tudelft.nl

Collin Sellman
Arizona State University
collin.sellman@asu.edu

ABSTRACT: The primary goal of this workshop is to produce open source data visualizations that help communicate results of learning analytics (LA) research to educators. Instructors are increasing their use of data to drive instruction, and various results of LA research are useful towards this end. However, the actionable insights discovered by the LA community are often inaccessible to educators due to their relative complexity. In this case, it is possible to use data visualization to communicate actionable insights about learners to optimize instruction effectively. Visualizations of learner data can make it easy for teachers and other education stakeholders to take evidence-based action. Organizers of the workshop intend to invite authors to describe and implement educational data visualizations that can aid decision making in online and offline classrooms. The workshop will result in a gallery of open source educational data visualizations that can be freely used by the LA community.

Keywords: Data Visualization, Data-Driven Instruction, Information Communication

1 INTRODUCTION

A recent survey in the United States found that 95% of the K-12 teachers use a combination of academic data and non-academic data to understand their students' performance. However, 34% of the surveyed teachers also reported that there was too much data for them to look at. How can we help educators make sense of large amounts of student data? Data visualization is one of the most widely used techniques that help people make sense of large amounts of numerical information. Graphical representations of data can be used very effectively to communicate context-specific information.

Reporting of learner data is one of the cornerstones of LA research, and the LA community has developed domain-specific data visualizations to show student learning in different contexts. Some of these visualizations emerged from the Learning Science research community (e.g., Learning Curves,) while other visualizations have a close affinity with classroom practice (e.g., Curriculum Pacing Plots.) Although these visualizations are slowly making their way into the hands of educators, many of these visualizations are not readily available for reproduction in the open source data analysis environments such and R and Python. This workshop aims to produce an open source gallery of education data visualizations that are easily reusable by LA researchers and practitioners.

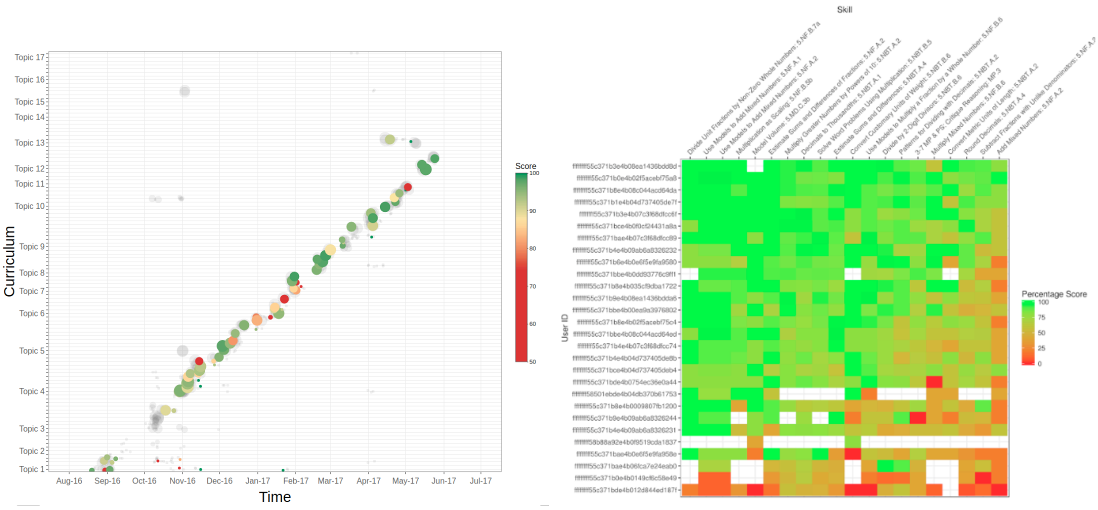


Figure 1a (left) and 1b (right): The figure on the left shows a Curriculum Pacing Plot, showing the progression of a classroom through a yearlong period. The figure on the right shows a Mastery Matrix for a classroom, which allows easy identification of struggling students and difficult topics.

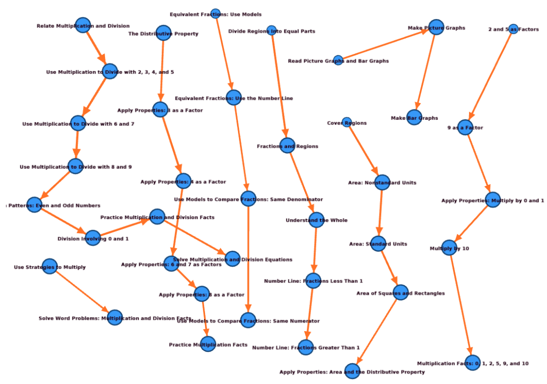


Figure 2: Learning Pathway Visualization, showing frequently taken learning paths taken by students in an online curriculum.

To make data visualizations a tool for knowledge discovery, they have to be made domain specific. A scatter plot, when adapted to a context of classroom, can become a Curriculum Pacing Plot (figure 1a,) a heatmap with student and skill data can be turned into a Mastery Matrix (figure 1b,) and a Learning Pathway Visualization with educational activities as nodes can display highways of student learning (figure 2.) All these visualizations are based upon commonly used graphical representations of data, but with a little tweaking, they turn into tools that let us extract meaningful information from educational data. Moreover, useful information visualizations can help support and engage a range of different stakeholders. For example, learner data visualizations can help curriculum coordinators in schools understand student behavior in a manner that can inform instruction. Learning Pathways of a MOOC can help instructional designer discover whether students are progressing through the curriculum as expected or not. Whereas many data mining techniques can produce “black boxes,” visualizations are human readable. This readability allows direct stakeholders (e.g., teachers and instructional designers) to critique assumptions made by data scientists, who are often removed from the context of data production.

LA research has the potential to inform real-world instruction, but bringing research findings into real-world requires effective communication to stakeholders outside the LA community. This can be difficult because educators usually find results of learning analytics challenging to understand. We believe that graphics can help us bridge this communication gap and make the results of LA research more fruitful.

2 ORGANIZATION

The workshop will be organized as a half-day event, as this is the first data visualization workshop being organized in the community. The organizers will invite authors to develop and implement open source data visualizations and describe them in posters, short papers, or full papers. Posters will describe visualizations that are relevant to the community but are still being developed. Short papers will describe works that are sufficiently mature but haven't been tested in the field, and full papers will describe novel visualizations that are mature and have been tested with the stakeholders in the field. The participation of the workshop will be mixed, and delegates other than the authors will be invited to register and take benefit of the workshop. The workshop will contain a series of visualization demos, and each presenter will show how everyone in the audience can use the open source visualization. Authors will be highly encouraged to develop their visualizations in R and Python, and all of the visualization programs will be uploaded in a GitHub repository that will remain freely accessible to the LA community. We expect 15 to 30 participants to attend our workshop. No special equipment other than a projector will be required.

3 OBJECTIVES

The goal of this workshop is to adapt existing visualizations or product novel visualizations of educational data that can communicate actionable information to educators. We will invite authors to describe and implement educational data visualizations relevant to a range of educational contexts such as various types of digital learning environments such as MOOCs, virtual schools, K-12 and university classrooms, and other exploratory learning environments like as educational games. We will encourage authors to produce interactive visualizations so that the users can actively engage with the data and use the graphs as tools to explore data and understand students rather than a snapshot of data that they can look at and reflect on. As the open-source code is an emphasis of the workshop, authors will also be asked to write programs that are easy to use, e.g., providing functions that take well-defined data structures (e.g., data tables, graphs) as input and produce desired visualizations as output. Topics for the data visualizations will be:

- Student learning and behavior
- Student knowledge and mastery
- Student learning trajectories and processes
- Student misconceptions
- Problem-solving strategies
- Teaching strategies
- Collaborative learning

- Classroom learning
- Emotional states
- Clickstream data
- Student groups and their differences
- Comparison of observed and ideal student behavior
- Usage and efficacy of educational content
- Demystifying “black box” machine learning models

Any other topics that are relevant to educational data and environments will also be included. We hope that the visualizations produced during this workshop can act as worked examples of visual design patterns that can be applied to educational data from a range of different sources and serve as a quick reference guide for LA community.

REFERENCES

Data Quality Campaign. (2018). *What Parents and Teachers Think About Educational Data.* Retrieved from https://dataqualitycampaign.org/resource/what-parents-and-teachers-think-about-education-data/

Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. Handbook of educational data mining, 43, 43-56.

Patel, N., Sharma, A., Sellman, C., & Lomas, D. (2018). Curriculum Pacing: A New Approach to Discover Instructional Practices in Classrooms. In *International Conference on Intelligent Tutoring Systems* (pp. 345-351). Springer, Cham.

Implications of Instructor Analytics Use Patterns for the Design of Actionable Educational Data Visualizations

Alyssa Friend Wise
New York University
alyssa.wise@nyu.edu

Yeonji Jung
New York University
yeonji.jung@nyu.edu

ABSTRACT: This paper offers insights to inform evidence-based learning analytics design through the presentation of an empirically-derived model of instructor analytics use. The model represents key elements of the ways in which instructors make sense of and respond to the analytics data as part of their pedagogical decision-making process, which can assist educational visualization designers in choosing among the myriad data representations possible to produce interpretable and actionable learning analytics systems. Instructor analytics use is shown as a multi-phase process divided across the larger activities of sense-making and pedagogical response. Sense-making process moves from a general area of curiosity which instructors can develop into more specific questions through interaction with the data, to reading the data to identify noteworthy patterns and appraising the patterns' importance in the course. Pedagogical responses involve taking the form of actions (whole-class scaffolding, targeted scaffolding, and revising course design) or adopting a wait-and-see holding pattern, and/or deeply reflecting on pedagogy. Drawing on this model, specific recommendations are made for how learning analytics design can align information presentation with core instructional practices and embed features to support processes of use in order to be most impactful on teaching and learning.

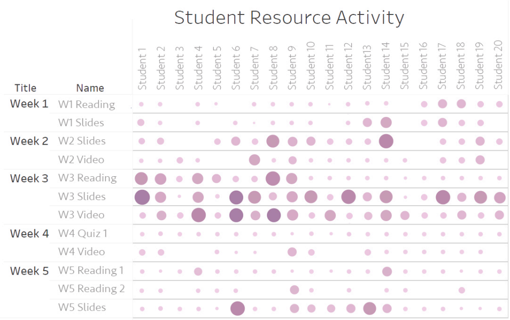
Keywords: Learning analytics use, Data-informed decision-making, Teaching analytics, Learning analytics design, Educational data visualization

1 INTRODUCTION

While the process of using analytics data to inform pedagogical decisions is acknowledged to be complex (Herodotou et al., 2017), little is known about the details of how it occurs in authentic teaching contexts. Data-informed pedagogical decision-making process involves more than just instructors' uptake of learning analytics tools; rather, it entails instructors' translation of tool-provided information into locally-meaningful knowledge and subsequently use of it to guide their pedagogical actions (Molenaar & van Campen, 2018). Examining such information use is critical to impact educational practice and inform the design of learning analytics in interpretable and actionable visualization. This paper fills a gap in the information available to educational visualization designers to make evidence-based design decisions by presenting an empirically-derived model of instructor analytics use to guide the design and implementation for learning analytics.

2 A BRIEF OVERVIEW OF THE MODEL DEVELOPMENT

The model was developed based on empirical case studies conducted with five university instructors who used a learning analytics dashboard in their teaching during the course of a semester. In-depth interviews were guided by the (limited) existing literature on the topic (Molenaar & van Campen, 2018; van Leeuwen, van Wermeskerken, Erkens, & Rummel, 2017; Verbert et al., 2013) and involved instructors' think-aloud walk-through of their analytics use, showing the relevant visualizations to concretize their responses. The dashboard used by these instructors was developed and rolled out by the university's Instructional Technology team based on consultations with each instructor about the kinds of student activity and performance information they would like to see for one of their courses. The personalized dashboard involved three to four distinct views (e.g. student access of course site and resources, video viewership, online quiz results or survey responses), each which acted as an independent overview into the data (see Figure 1 and 2). An inductive qualitative analysis was conducted on the interview transcripts and surfaced twenty emergent themes related to instructors' practices of analytics use. The themes were then synthesized into a situated model of instructor analytics use that is presented in the next section. The full study, including details and evidence for each theme and the model development process, is reported in Wise and Jung (in review).



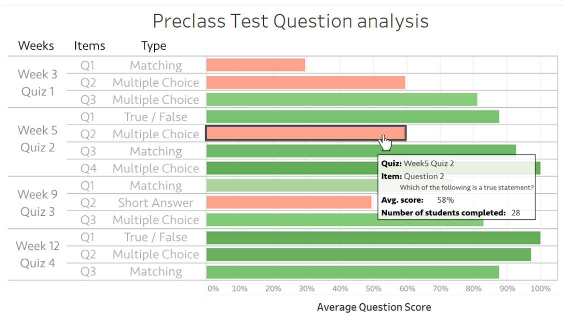


Figure 1. Example Dashboard View of Resource Access **Figure 2. Example Dashboard View of Quiz Results**

3 A SITUATED MODEL OF INSTRUCTOR ANALYTICS USE

The model consists of multiple phases of activity embedded in a two-part structure with sense-making and pedagogical response as distinct aspects of practices (see Figure 3). Such structure aligns with the majority of prior studies which have described instructor analytics use as first determining an understanding of what the information tells about the course and then considering potential actions in response to it (Herodotou et al., 2017; van Leeuwen et al., 2015).

Looking inside each part of the process, sense-making begins with an instructor's general area of curiosity (e.g. class-level or individual-level engagement, usefulness of course materials). While prior studies suggest that instructors can come to analytics with fully-formed questions (Dyckhoff et al., 2012) either based on prior analytics use or their own methods of data collection and analysis, or that they may just respond to the data as presented (Herodotou et al., 2017), this model offers a third possible path: that instructors may start their analytics use with a general area of curiosity. Areas of curiosity can develop into more specific questions through interaction with the data, which can then be answered with more careful examination (e.g. identifying potential relationships in the data that instructors hope to explore further) (Molenaar & van Campen, 2018).

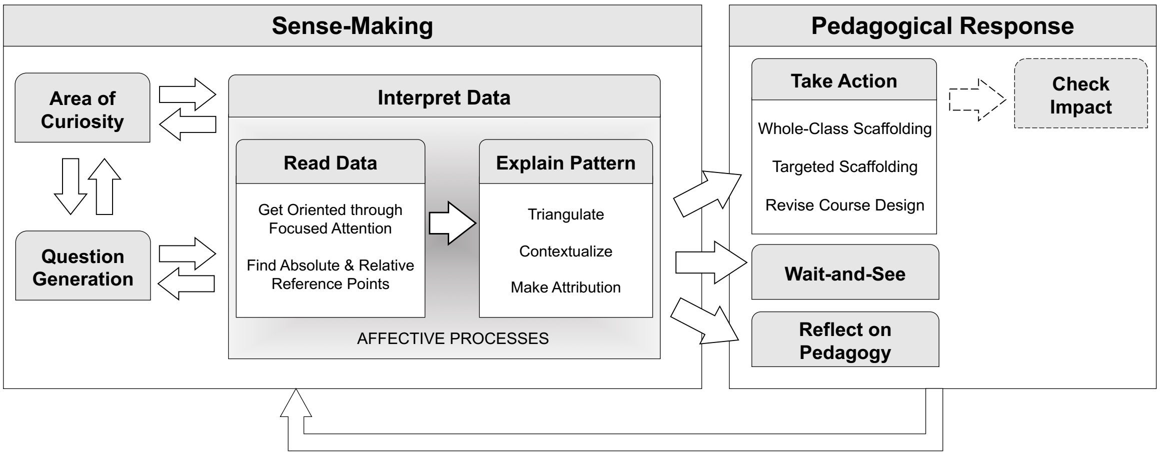


Figure 3. A Situated Model of Instructor Analytics Use

Data interpretation occurs through two distinct and equally important phases of activity: reading the data to identify meaningful patterns and then generating explanations that address the patterns' importance for the course. In reading the data, instructors get oriented to visualizations through paying attention to a specific piece of information or a noticeable pattern as initial anchors, and then expanding their view outwards to explore the different kinds of information offered by the overall analytics. This differs from previous studies which suggested a sequence of first getting oriented, then applying focused attention to specific data (Van Leeuwen et al., 2017). Such difference may depend on how the analytics view is organized as certain kinds of layout design and information arrangement can allow instructors to get oriented first; thus both activity sequences should be considered as possibilities. In reading the data, there is a need for reference points of comparison which instructors can feel confused about what to use as; either absolute (e.g. do students engage with at least 75% of the provided materials) or relative (e.g. does student engagement change during the course, how does this year's pacing trajectory compare to last year's). Consideration of how to best provide explicit reference points to aid users in navigating through data is a growing area of attention in the literature (for example see Patel, Sharma, Sellman, & Lomas, 2018). In explaining patterns, instructors extend the meaning of patterns identified by explaining (or questioning) their implications as related to their course. Instructors often try to triangulate the patterns with additional information (e.g. class observation) to confirm their interpretation. When this supports the interpretations, instructors may use their contextual knowledge of the course and students to explain what the results might mean and make attributions to potential causes (Molenaar & van Campen, 2018). When the external information and analytics data do not align, it can lead instructors to question the analytics (Dazo, Stepanek, Chauhan, & Dorn, 2017) and/or hesitate to take action (Herodotou et al., 2017). In addition to cognitive processing of patterns, data interpretation can provoke affective responses such as surprise, disappointment, or joy as reported in the literature (Wise, Zhao, & Hausknecht, 2014).

Following sense-making, pedagogical response is instructors' decisions and/or changes in thinking based on the analytics. The most common response type is taking action towards (1) the whole-class, (2) particular students, or (3) course materials (e.g. Herodotou et al., 2017; Molenaar & van Campen, 2018; Xhakaj, Aleven, & McLaren, 2017). Checking whether the actions taken have achieved the intended impact is a final phase to close the loop; however this does not always occur (Dazo et al., 2017). Another common response to analytics is to adopt a holding pattern of waiting to see what will happen as more data is made available (Herodotou et al., 2017). Deep reflection and shifts in how an instructor conceptualizes their pedagogy is a new, interesting response type that has not received much attention in the literature (c.f. Molenaar & van Campen, 2018) but may have greater and longer-lasting effects than simple adjustments to teaching.

In addition to the unidirectional path through the different phases described above, instructor analytics use may also occur iteratively within and across each of the two larger parts (e.g. access to data leads to new questions; actions taken to test an initial interpretation influence back to the interpretation itself).

4 IMPLICATIONS FOR LEARNING ANALYTICS DESIGN

This situated model offers a clear starting place for efforts to design learning analytics to support instructors' pedagogical decision-making practices, which can guide designers in thinking ahead to instructors' analytics use during the design process (Xhakaj et al., 2017). In making evidence-based design decisions, it is critical to work directly with educational stakeholders throughout the development process. This process, however, requires more than just asking stakeholders what information they would like to look at and use, since it is a difficult question to answer in the absence of prior experience working with such data. More details should be considered in this process, including how gaining access to the data contributes to shifts in instructors' understanding or new question generated to the data. This highlights the need for evidence-based design to be attentive to actionability where "analytics connect with education and the changes that administrators, teachers and students want the tools to make in order to support their everyday learning, teaching and assessment work" (Ferguson et al., 2016, p. 9). The core areas for attention are presented in the following sets of design recommendations based on the situated model.

4.1 Learning analytics design should align information structures with instructors' pedagogical concerns.

Organize information from the perspective of instructors, not data structures. In analytics development, it is easy (and often necessary) to start thinking about the data in the form in which it is made available (e.g. organized alphabetically, by type of interaction, or by system-time). But instructors often think in different categories: weeks or units of a class, sets of associated course activities. This disconnection creates a critical barrier for effective analytics use, but can be addressed by explicitly eliciting instructor conceptualizations of how they think about their course and the different elements that compose it (which is a quite different kind of question to ask instructors than what kinds of information they would like to know). Attention to this issue may also raise the need for (re)considering learning design before analytics are built so that the two can be in alignment (Lockyer et al., 2013).

Align the timing of system and instructors' practices. Similar to the prior consideration on the organization of information, the timing of access to information needs to be considered from the perspective of the instructor. The deferred update of data refresh can limit the usefulness of the analytics for instructors who want to access the dashboard immediately prior to a class as reported in Wise and Jung (in review). While constant data updating for the entire system may not be realistic, allowing instructors to update their data on demand in situations of need could be one way to address the situation.

4.2 Learning analytics should be designed to support processes of use.

Embed support for question generation and maintenance. A key element of the value proposition for instructor analytics use is that analytics data responds to important questions on which instructors can take action. As an iterative process within and across each of the two larger parts of the model, questions often emerge or are refined through further examination of the data. Importantly, amidst hectic teaching schedules such questions may not be formed or remembered across analytics use sessions. Analytics tools can be designed to support this process by including features which support the generation and maintenance of questions (and perhaps answers). For example, a question area associated with a visualization could offer a set of editable, tailorable questions with add, delete and edit functions that provide flexibility of use. Questions can be maintained across sessions of use and annotated with answers instructors find in the data or tags for future follow-up.

Incorporate visual aids to find entry points to the analytics. Another important consideration is how to facilitate instructors to find entry points with which they can get oriented to the analytics. Rather than beginning with an overview and then digging in a particular part, instructors may often begin with some part of the analytics that they can make sense of and expand it outwards (Wise & Jung, in review). Visual aids can be incorporated into analytics to support this process; for example, in a large matrix of data toggleable tools which highlight information by rows or columns could help users focus their attention on finding certain kinds of patterns or extended sequences of visual information which can allow grouping to let users attend to individual weeks or quizzes in turn.

Help instructors to find and work with reference points for data interpretation. A further consideration to support instructors in making use of analytics is offering support for finding reference points with which to make sense of the data. Providing access to similar data from prior terms or overarching trends from similar courses or tools for making comparisons across time can provide high level relative reference points. Absolute reference points may be explicitly elicited through a process of guided reflection through which instructors articulate their expectations for class activity, engagement or performance in terms of the metrics available.

Embed flags for later decisions to take action and check impact. One of the pedagogical responses that instructors can commonly make is wait-and-see approach that delays action till further data is available. Effectiveness of this strategy requires the instructor to remember the situation they are waiting to know more about and return to reexamine it at some future date. However, it is quite possible that this never occurs. Rather than relying on instructor memory, analytics design can support this process by offering features that let instructors mark and/or annotate a pattern they observe in the data for future follow-up. Similarly, when action is taken, analytics features can be used to create externalized reminders to check the impact of the action.

4.3 Learning analytics should be designed to support sharing and conversations.

Help instructors to share the analytics by offering de-identified views. When instructors make sense of the analytics or take actions based on it, instructors may need or want to share analytics with other instructors to engage in a process of collaborative interpretation or with students as an object for discussion and reflection in the class. However, this practice may raise potential privacy concerns regarding the use of analytics-as-evidence. One way to facilitate the use of analytics as a mediational object (Wise, 2014) is to make it easy for instructors to switch to a view in which student identities have been removed or hidden.

5 IMPLICATIONS FOR LEARNING ANALYTICS IMPLEMENTATION

In addition to rethinking learning analytics design in the context of instructional practices, it is also important to consider ways in which analytics use can be facilitated through pedagogical support for the process of use itself. This is critical to facilitate instructors' translation of information into actionable insights which can feed back to the evidence-based design processes. One potential way is to consider ways to educate instructors in how to work with data to inform their teaching. This can be done upfront through structured instructor data training or in-situ with the introduction of a pedagogical analytics coach who creates a series of scaffolds to support instructors in this translation process. Analytics coach supports can take the form of periodic emails, one-on-one coaching sessions, department-based workshops and the cultivation of local instructor communities. For example, email messages can highlight particular pedagogical questions (e.g. how can I find and help students who seem unengaged in the first few weeks?), explaining how to use the dashboard to answer it (e.g. open the student-course interaction grid and look across the rows for consistently light colored cells), and then discussing actions that could be taken in response (e.g. speak with them individually to find out what is going on and make them aware that you are invested in their success, highlight for the whole class habits of successful students). In this way, pedagogically meaningful questions, answers, and actions are linked together to frame analytics use, rather than starting with a data-centric orientation. The same issues can be engaged with on a broader scale through one-on-one coaching or workshops in which instructors are supported in working through these sequences using data from their own classes. Collaborative interpretation with a sample data set can be also implemented to discuss common challenges in analytics use and disambiguate the meaning of analytics through dialogue with other participants. In the long term, local instructor communities of practice around analytics use can be cultivated where such contextualized, embedded, ongoing support networks are more effective than short-term information delivery (Darling-Hammond & Richardson, 2009). Put together, the sets of recommendations highlight the importance of establishing opportunities for future efforts in both learning analytics design and implementation based on empirical findings for supporting instructors' situated use of analytics.

6 CONCLUSION

This paper offers insights to inform evidence-based learning analytics design decisions through the presentation of an empirically-derived model of instructor analytics use. An understanding of the practices instructors engage in when using analytics in their teaching can guide educational visualization designers in choosing among the myriad data representations possible to produce interpretable and actionable learning analytics systems. Future work can investigate the impact of the decisions taken to validate or refine the recommendations made above. In addition, broader data collection on instructor analytics use including log-file records, experience sampling data, and classroom observations can further reveal how analytics use occurs and feeds back into instructors' teaching practices. Together these efforts help strengthen the lines of communication between stakeholders and designers, and help us more as a field towards evidence-based learning analytics design as a practice to support teaching and learning, fostering educational success.

REFERENCES

Darling-Hammond, L., & Richardson, N. (2009). Research review/teacher learning: What matters. *Educational leadership, 66*(5), 46-53.

Dazo, S. L., Stepanek, N. R., Chauhan, A., & Dorn, B. (2017, May). Examining instructor use of learning analytics. In Proceedings of *the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2504-2510). New York, NY: ACM.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society, 15*(3), 58–76.

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittlemeier, J., Rienties, B., Ullman, T., & Vuorikari, R. (2016). Research evidence on the use of learning analytics: Implications for education policy. In R. Vuorikari, & J. Castano-Munoz (Eds.), *A European framework for action on learning analytics*. Luxembourg: Joint Research Centre Science for Policy Report.

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., Hlosta, M., & Naydenova, G. (2017, March). Implementing predictive learning analytics on a large scale: the teacher's perspective. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 267-271). New York, NY: ACM.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439-1459.

Molenaar, I., & van Campen, C. (2018). How teachers make dashboard information actionable. *IEEE Transactions on Learning Technologies*.

Patel, N., Sharma, A., Sellman, C., & Lomas, D. (2018, June). Curriculum Pacing: A New Approach to Discover Instructional Practices in Classrooms. In *Proceedings of International Conference on Intelligent Tutoring Systems* (pp. 345-351). Cham, Switzerland: Springer.

van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education, 90*, 80-94.

van Leeuwen, A., van Wermeskerken, M., Erkens, G., & Rummel, N. (2017). Measuring teacher sense making strategies of learning analytics: a case study. *Learning: Research and Practice, 3*(1), 42-58.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500-1509.

Wise, A. F. (2014, March). Designing pedagogical interventions to support student use of learning analytics. In *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 203-211). New York, NY: ACM.

Wise, A. F., & Jung, Y. (in review). Teaching with analytics: Towards a situated model of instructional decision-making.

Wise, A. F., Zhao, Y., & Hausknecht, S. (2014). Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics, 1*(2), 48-71.

Xhakaj, F., Aleven, V., & McLaren, B. M. (2017, September). Effects of a teacher dashboard for an intelligent tutoring system on teacher knowledge, lesson planning, lessons and student learning. In *Proceedings of the 2017 International Conference on Artificial Intelligence in Education* (pp. 315-329). Cham, Switzerland: Springer.

Visualizing the Solution Space of Educational Games using TRESTLE

Erik Harpstead
Carnegie Mellon University
harpstead@cmu.edu

Christopher J. MacLellan
Soar Technology, Inc.
chris.maclellan@soartech.com

ABSTRACT: When designing open-ended educational games and other creative instructional environments it is important for designers to understand what learners can do within the space their games afford and whether behaviors across that space are supporting their instructional goals. In this demo we will present several prototype visualization concepts based on the TRESTLE concept formation algorithm to organize data from student solutions into a tree structure amenable to several kinds of visualization.

Keywords: Visualization, concept formation, alignment, solution space

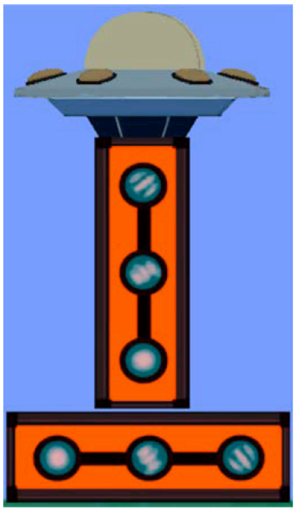
1 INTRODUCTION

Exploratory data analysis is an important step within the educational data mining process, particularly so in the context of educational games, which generate large amounts player data. Being able to visually inspect trends in data can provide context and perspective on complex statistical analyses and can help guide educational technology design. Unfortunately, hand conducting such analyses on larger data sets is often too unwieldy and time consuming to be practical. To overcome this barrier, we developed the TRESTLE algorithm and an accompanying set of visualizations to help designers and researchers hierarchically organize and explore structured data, such as the kind generated by educational games and other open-ended, creative instructional environments.

Understanding the breadth of approaches that learners can take in these environments and, more importantly, how the game reacts to those approaches, is essential for ensuring effective instruction. In this paper we briefly describe the TRESTLE approach and describe two examples of how it can support the organization, exploration, and interpretation of structured educational game data.

2 TRESTLE

TRESTLE is a concept formation algorithm that incrementally learns conceptual hierarchies given structured examples as training data (MacLellan, Harpstead, Aleven, & Koedinger, 2016). Unlike most learning systems that only support a vector of feature values, TRESTLE supports hierarchical attribute-value lists (represented as Python dictionaries) that contain both nominal (e.g., discrete object types) and numeric (e.g., x and y position) attributes as well as attributes that refer to nested attribute-value lists, which we refer to as structural attributes (e.g., "block1" might have a nested set of attribute-values that describes its location and type). It also supports relational attributes that can describe relationships between other attributes, such as specifying that "block1" is on top of "block2", e.g., on(block1, block2)[1]. The variety of attribute types that TRESTLE can handle makes it broadly applicable to wide range of potential data sets.



```
{ _Success: "False",
  ?Block1: {type: "ufo",
       angle:0.0,
       left:0.1
       right:2.8,
       bottom:4.1,
       top:5.6},
  ?Block2: {type: "rectangle",
       angle:90.0,
       left:0.9
       right:1.9,
       bottom:1.1,
       top:4.1},
  ?Block3: {type: "rectangle",
       angle:0.0,
       left:0.0
       right:3.0,
       bottom:0.0,
       top:1.0},
  (On ?Block1 ?Block2): "True",
  (On ?Block2 ?Block3): "True"}
```

Figure 1. An example game state from *RumbleBlocks* and how it would be described to TRESTLE.

Given structured examples described in this representation, TRESTLE can engage in both supervised and unsupervised learning or a combination of the two. Specifically, the system can learn a shared hierarchical organization of both labeled and unlabeled data that enables learning from one kind of data to benefit the other kind. Learning within TRESTLE is incremental, meaning that it is presented with a sequence of examples. Upon receiving each example, TRESTLE sorts each new example into its hierarchy, updating it to reflect the new training data. To guide this learning process TRESTLE uses an objective function called category utility, which is derived from psychological studies of human concept formation (Fisher, 1987). This objective function is similar to the information-gain metric used in decision tree learning but supports the ability to predict arbitrary attributes of examples.

The hierarchical knowledge structure learned by TRESTLE supports two key capabilities: prediction and clustering. Prediction within TRESTLE operates similarly to learning. The system accepts as input examples with some of their attribute values missing. The system sorts these partial examples into its current organization using the available features and the resulting cluster it is sorted into is used to predict the values of any missing attributes. In this regard, TRESTLE is similar to other instance-based learning approaches, such as k-nearest neighbor, but it automatically determines—based on the data—how many examples (the k) to use for prediction. Previous work suggests that TRESTLE's prediction performance is similar to humans on the task of labeling the stability of block structures generated by students in an educational game (MacLellan et al., 2016).

[1] Additional specifics on the semantics of this instance representation are available from TRESTLE's online documentation: https://concept-formation.readthedocs.io/en/latest/instance_representation.html

In addition to prediction, TRESTLE can cluster examples both hierarchically and into flat clusters (e.g., into k groups). To generate clusterings, users present TRESTLE with a sequence of complete or partial examples (labeled or unlabeled), which it organizes into a conceptual hierarchy using its learning mechanisms. The resulting hierarchy can be directly returned as a clustering of the data. Additionally, TRESTLE has multiple post-processing routines that can translate these hierarchies into flat clusterings of the examples. For example, it can start at the root of the hierarchy (which contains all examples) and progressively break this top-level clustering into progressively smaller and smaller groups, stopping when it has optimized one of a range of metrics, such as Category Utility, AIC, or BIC. Our analysis of TRESTLE's ability to cluster block structures suggests that it produces groupings of examples that have reasonably high agreement with human-generated clusterings of the same blocks structures (MacLellan et al., 2018), which suggests that TRESTLE might be used to organize large volumes of examples in a way that aligns with how humans would organize the same data.

3 VISUALIZATION USE CASES

We have designed several visualizations to facilitate interpretation of TRESTLE's outputs, though few have been empirically validated with target users. These visualizations have been built to be interactive using D3.js (Bostock, Ogievetsky, & Heer, 2011) in order to enable an analyst to explore the information and make their own reflective judgements about their design.

Each of the visualizations presented here was developed as part of the analysis of *RumbleBlocks* (Christel et al., 2012) an educational game designed to teach based concepts of structural stability and balance to young children. In this game players build block towers that must survive an earthquake to succeed. Given that the game relied on a real time physics engine to generate its feedback it was not always clear if the game was providing clear guidance to players that would help them learn its targeted concepts.

3.1 TRESTLE Tree Visualization

The core visualization of TRESTLE is a visual representation of its hierarchical concept tree. Figure 1 shows this visualization of examples of player solutions to one level in *RumbleBlocks*. In this visualization each circle represents a collection of student solutions to a game level. The leaf concepts (filled in circles) represent specific instances within the dataset while the transparent enclosing circles represent higher-order clusters containing subgroups. The size of each cluster represents how many instances are grouped within it. To the right of the tree is a control panel showing various options as well as an Attribute-Value Table that shows the distributions of each attribute-value within the tree. When a concept or instance is clicked on, the view zooms to focus on that concept and the Attribute-Value Table updates shows the distribution of properties within the selected concept/instance.

Nodes in the visualization can be colored according to their different attributes to highlight trends within the concept tree. In main example in Figure 1, the nodes are colored based on whether a solution succeeded or failed the level in question according to the game's log data. Solutions are colored yellow if they are more likely to pass the level, and purple if they are more likely to fail. In this case there is a clear successful cluster (bottom of left branch), and a mostly clear negative cluster (right branch). The outcomes of the other two main branches of the tree (left and right of the larger left branch) are less clear and would potentially warrant further investigation. An analyst can re-color the same visualization according to different properties of solutions to see if there are any correlations in trends that might warrant investigation as design issues. The smaller examples in Figure 1 are each colored according to a different physical property of the solution tower (e.g., symmetry, base width, center of mass height).In principle these properties should roughly correspond with succeeding at the level but there is not a clear correspondence to the coloring based on success, suggesting there might an issue with how feedback is assigned in the game.

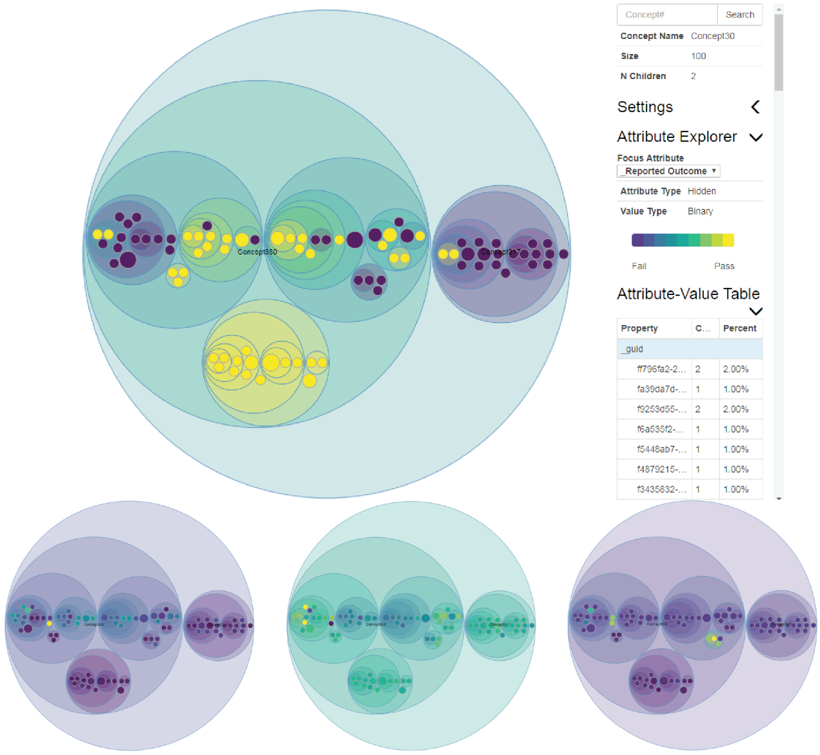


Figure 2. A visualized TRESTLE tree of a sample of 100 solutions to a level in RumbleBlocks. In the top visualization clusters are colored by their likelihood of succeeding on the level (yellow for passing, purple for failing). In the lower three visualizations the same tree is re-colored according to different physical properties of game solutions.

3.2 TRESTLE Alignment Visualization

While the hierarchical tree is currently the default output for visualization in TRESTLE, we have also developed more advanced forms of visualization that make use of the TRESTLE data structure in a flattened form. The alignment visualization shown in Figure 3 is based on analyses we have done of the solution space of *RumbleBlocks* (Harpstead, MacLellan, Aleven, & Myers, 2014). This visualization breaks up the TRESTLE tree into representative clusters that can be plotted according to their properties to look for correlations in data that might be indicative of problems. In the case shown in Figure 3 a principle relevant metric (in this case the symmetry of the tower) should be predictive of successful performance in the game. within the visualization this is denoted by the red and green shaded regions where it would be desirable for solution clusters to fall in the green areas while it would be a problem if they mainly fell into red regions.

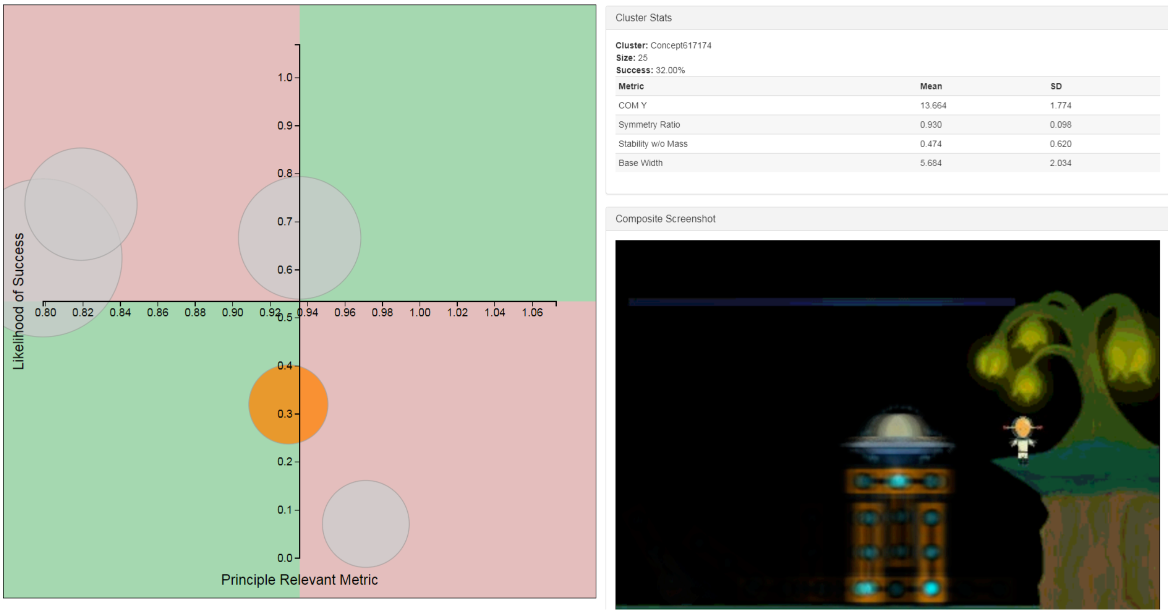


Figure 3. An example of employing TRESTLE to visualize the alignment of clusters of player solutions to a game. Solutions are plotted along an axis for principle relevant metrics (x-axis) and likelihood of success (y-axis). Ideally there would be a relationship between principle relevant metrics and success, denoted by the shaded areas. [2]

To support digging further into apparent trends the visualization supports the option of including screenshots that can be linked to instances within the clustering. These screenshots can be composited together (lower right of Figure 3) to allow an analyst to more quickly examine why a particular trend might be happening within their instructional environment and what actions they may explore to fix the problem.

4 CONCLUSION

Our goal in designing visualizations for TRESTLE is to help analysts to organize their data in way that can support intuitive exploration. We have found this to be particularly useful within datasets from complex instructional environments such as educational games. We hope that others can see utility in this approach and can find a way to apply it to their own contexts.

REFERENCES

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D[3] Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

Christel, M. G., Stevens, S. M., Maher, B. S., Brice, S., Champer, M., Jayapalan, L., … Lomas, D. (2012). RumbleBlocks: Teaching science concepts to young children through a unity game. In *Proc CGAMES 2012* (pp. 162–166). IEEE. https://doi.org/10.1109/CGames.2012.6314570

Fisher, D. H. (1987). Knowledge Acquisition Via Incremental Conceptual Clustering Learning ~ Element Performance Element. *Machine Learning*, *2*, 139–172.

[2] A live version of this visualization is available at http://erikharpstead.net/alignment/visualization.html

Harpstead, E., MacLellan, C. J., Aleven, V., & Myers, B. A. (2014). Using extracted features to inform alignment-driven design ideas in an educational game. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 3329–3338). New York, New York, USA: ACM Press. https://doi.org/10.1145/2556288.2557393

MacLellan, C. J., Harpstead, E., Aleven, V., & Koedinger, K. R. (2016). TRESTLE: A Model of Concept Formation in Structured Domains. *Advances in Cognitive Systems*, *4*, 131–150. Retrieved from http://www.cogsys.org/papers/ACSvol4/paper10.pdf

"What do students know, how long does it take them to know?" at a Glance for Teachers and Instructional Designers

Truong-Sinh An, Majd Edriss, Agathe Merceron, Thuy-Anh Nguyen
Beuth University of Applied Sciences
truong-sinh.an@outlook.com, edriss.majd@gmail.com, merceron@beuth-hochschule.de, thuyanh.nguyen.de@gmail.com

ABSTRACT: In this paper, we present the visualizations realized in the learning analytics service for the Learning Companion Application. User interactions stored by the application form the evidence for these visualizations. The diagrams for teachers enable them to grasp at a glance which topics their students master or not so that they can prepare their next class accordingly. The same diagram offers additional options, like the total number of attempts for instructional designers so that they can reflect on the difficulty level of the exercises. An additional visualization for instructional designers shows the time students spend on each learning object. The LA service is being realized as an LTI-Tool.

Keywords: traffic-light diagrams, xAPI statements, learning locker, elasticsearch, grafana

1 INTRODUCTION

The Learning Companion Application (LCA) is developed in the smart learning project[1] to fit the needs of full-time employees who take part in an Energy Consultant training in a Chamber of Crafts. LCA can be thought of as a learning management system (LMS) with two distinctive features. First, the digital learning resources are stored centrally in a repository and can be accessed without replication when a course is taught in different institutions. Second, it includes a recommendation service for learners which selects appropriate contents, as well as a learning analytics service to different stakeholders, in particular to teachers and instructional designers. LCA is independent of any topic and any institution and, therefore, can be used in other contexts and for other courses (Krauss et al. 2017).

A course in LCA as in many LMS can be divided into sections, which can be divided into learning units. A learning unit contains different learning resources also called learning objects (LO) such as texts, videos, animations, PDF files, other media-types, and exercises. These learning objects can be reused in other courses. To support the pedagogical concept adopted in LCA as well as to implement the recommendation and learning analytics services, metadata are associated with any learning object. These metadata contain among others at least one learning objective and a typical learning time. A learning unit is rendered as an accordion with a specific sequential structure, see Figure 1. The top item is the list of the learning objectives of all learning objects of that unit. A learner can rate each learning objective and so reflect on how much s/he knows already on that topic, from 1

[1] https://projekt.beuth-hochschule.de/smartlearning/

(know nothing) to 5 (expert). This item is followed by the sequence of the LOs of that unit. In Figure 1, this list includes a set of exercises shown in orange. Following all LOs, the next item in the accordion-view is again the list of learning objectives. By rating them, a student can reflect on how much s/he knows after learning the unit. The follower item allows students to provide feedback on the typical learning time for that unit (from 1, way too little time to 5, way too much) and give comments. The last item in the list opens a discussion thread on that unit. These last two items are marked a Communication-tools in Figure 1.

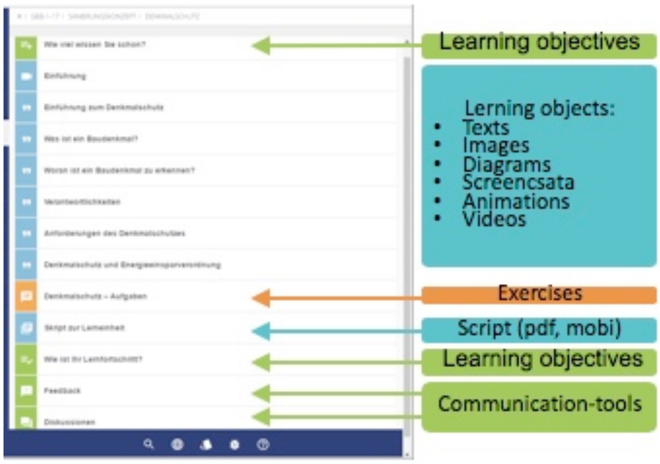


Figure 1: A learning unit in LCA

The first aim of the learning analytics (LA) service is to support teachers and instructional designers. During the project, three meetings with three teachers (N=3) of the Chamber of Crafts have taken place to sense their needs and to discuss proposed solutions. The outcome stressed the importance of a simple and unambiguous visualization: teachers should clearly understand what they see and not be overwhelmed with too much. Therefore, we have opted for well-known diagrams that teachers are familiar with. The LA service should enable teachers to be aware of how many students are mastering, are in the process of mastering or do not master at all the topics of a learning unit, so that they can prepare their next class according to the learning needs of their students. The LA service should enable instructional designers to improve the learning objects they create in cooperation with the teachers. For them too, it is important to understand what they see. However, they may need to explore more students' interactions to be aware of whether the resources they develop have the right length or the right level of difficulty.

In the next section, we describe the interactions data stored by the system. In the follower section, the diagrams for teachers and instructional designers are presented. This paper ends with a conclusion and future works.

2 DATA AND TOOLS

Comprehensive user interactions are stored as xAPI[2] statements in Learning Locker[3]. Examples of stored interactions include the opening of a learning unit, opening and closing of every single learning object, self-assessment of each learning objective, attempt in solving an exercise, starting, pausing or quitting a video etc. As an example, consider the following xAPI statement:


```
{ "actor": { "mbox_sha1sum": "13648454125cf6ef31a9e632389c9a806316c9ad" }, (1)
  "verb": { "id": "http://adlnet.gov/expapi/verbs/answered" }, (2)
  "object": { (3)
    "id": "https://vfh143.beuth-hochschule.de/…?itemID=U05LX0ZUU19BRkdfRmV1Y2h0ZXNjaHV0el8wMV9NQw",
    "definition": { "type": "http://adlnet.gov/expapi/activities/question",
      "name": { "de-DE": "Bauphysikalische Grundlagen" } }
  },
  "result": { (4)
    "score": { "scaled": 0.5, "min": -1, "max": 1 },
    "response": "[\"Die Wasserdampfsättigungsmenge ist die Höchstmenge an Wasserdampf die Luft bei
einer bestimmten Temperatur aufnahmen kann.\"]",
    "duration": "PT0H1M11S",
    "extensions": {
      "https://slehwr&46;beuth-hochschule&46;de/xapi/extensions/questionType": "choiceMultiple",
      "https://slehwr&46;beuth-hochschule&46;de/xapi/extensions/correctResponsePattern": [
        "Die Wasserdampfsättigungsmenge ist die Höchstmenge an Wasserdampf die Luft bei einer
bestimmten Temperatur aufnahmen kann.",
        "Der Wasserdampfdruck ist abhängig von der relativen Luftfeuchtigkeit und der Lufttemperatur."
      ]
    }
  },
  "context": { (5)
    "platform": "moodle.hwk-berlin.de", (6)
    "statement": { "id": "db36072e-6759-401a-bc7b-daad0677b683" }, (7)
    "contextActivities": { (8)
      "parent": [
        { "id": "https://vfh143.beuth-hochschule.de/…?itemID=U05LX0ZUU19BRkdfRmV1Y2h0ZXNjaHV0eg",
          "definition": { "type": "http://adlnet.gov/expapi/activities/interaction" } }
      ],
      "grouping": [
        { "id": "https://vfh143.beuth-hochschule.de/api/lcms/courses/GEB",
          "definition": { "type": "http://adlnet.gov/expapi/activities/course" }
        }, { "id": "https://vfh143.beuth-hochschule.de/…?itemID=U05LX0ZUUw==",
          "definition": { "type": "http://adlnet.gov/expapi/activities/module" }
        }, { "id": "https://vfh143.beuth-hochschule.de/…?itemID=U05LX0ZUU19BRkdfRmV1Y2h0ZXNjaHV0eg",
          "definition": { "type": "http://adlnet.gov/expapi/activities/interaction" } }
      ]
    },
    "extensions": {
      "http://adlnet&46;gov/expapi/activities/course": [ "GEB-1-17#GEB" ] (9)
    }
  },
  "timestamp": "2017-03-30T13:30:15.152500+00:00", (10)
  "id": "01a785e0-d77f-4268-860b-2b32883d6c7e" (11) }
```


The xAPI statement with the given *id* **(11)** above contains the information that a *specific actor* **(1)** did *answer* **(2)** a *specific question* **(3)** with the shown *result* **(4)** on a specific *timestamp* **(10)**. The *scaled*

---

[2] https://github.com/adlnet/xAPI-Spec

[3] https://github.com/LearningLocker/learninglocker

score of 0.5 indicates that the given solution is partially correct; the question was displayed for a *duration* of 1 minute and 11 seconds. For further analysis, the given *response* and the *correct response pattern* are also stored. For the purpose of better understanding the data, further information is bundled in the *context* **(5)**. The statement reference **(7)** links to the prior stored xAPI statement on a higher level which allows to build a graph of the learning behavior; xAPI statements on the same level, e.g. multiple attempts of *answering* the same question, will refer to the same higher statement **(7)** within this learning session – their *timestamps* **(10)** help to order the attempts. Information about the parent **(8),** like the exercise this question belongs to, and grouping helps to distinguish if a learning object, here the given question **(3)**, is used in several learning units or courses. As the same course can run several times, the *platform* of the host institution **(6)** and the internal *course short name/id* in that platform **(9)** help to distinguish between the instances.

The visualizations are realized in the LA service as a plug-in of the Grafana[4] framework. We use the *statement-forwarding* feature of learning locker since version 2 to sync the statements with elasticsearch[5] from which Grafana reads the data. Initial import is realized using an own tool.

3 VISUALIZATION

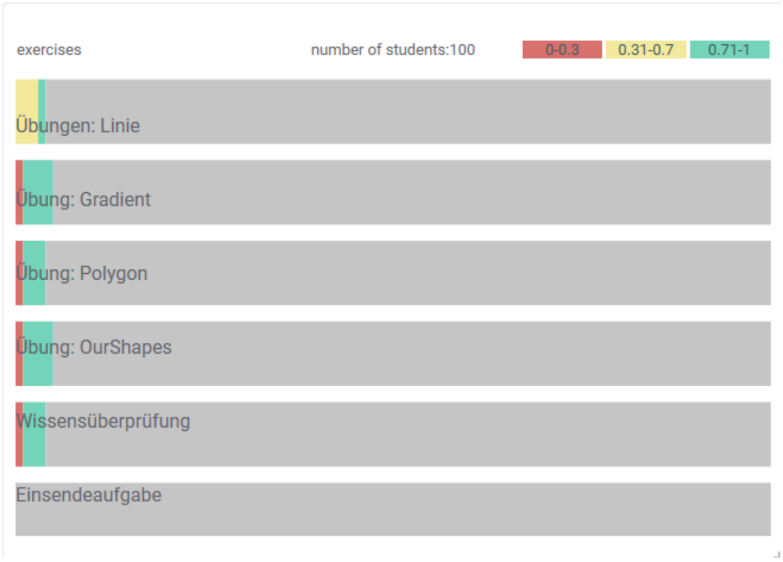


Figure 2: Many of the 100 students did not attempt any exercise (long grey bars). The exercises Gradient and OurShapes were attempted the most and mostly correctly solved (large green area)

The visualization depicted in Figure 2 enables teachers to grasp at a glance the performance level of their students at the level of a learning unit. It uses the well-known traffic-light metaphor used in other works as well, for example in (Dollár & Steif 2012). Teachers see for each exercise of the unit how many students are in green – correct solution –, yellow – solution partially correct –, red – wrong solution –, or grey – no attempt. The time span and the threshold values from red to yellow

[4] https://github.com/grafana/grafana

[5] https://github.com/elastic/elasticsearch

and from yellow to green can be chosen by the user. Figure 1, top corner right, shows the threshold values 0.7 and 0.3: if an exercise has got between 31% and 70% of the maximal score, it counts as partially correct and is color-coded in yellow. There are several options to calculate the performance of a student on an exercise. The default value set for teachers is simply the score of the last attempt as it reflects the best the current knowledge of students so that teachers can adjust their next lesson accordingly. Other metrics are available and can be chosen in a drop-down list. They are primarily for instructional designers to give them awareness on how difficult or easy it was for the students to solve that exercise. As an action, instructional designers in cooperation with teachers might rework that exercise to make it easier or more difficult to solve, or, leave it as is. These metrics are the average, minimal and maximal score on all attempts, as well as the total number of attempts. A general large number of attempts might indicate that the difficulty level is not appropriate.

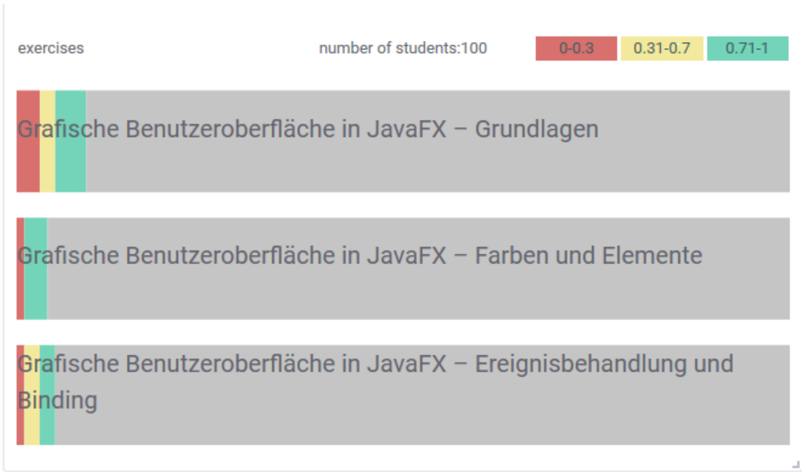


Figure 3: Overview of three learning units

Discussions with teachers have shown that they also need a general overview to plan remediation classes like at the end of a course to tackle again questions that have not been well understood by their students. They need an overview at the course level; if they detect an important part of students in the red or yellow area, they might want to spot the problematic topics, drilling down at the section level, then at the learning unit level as shown in Figure 3, and then into the unit itself and get the visualization presented in Figure 2. For this situation, we have developed a similar visualization: green, yellow, red and grey. Starting from the visualization depicted in Figure 2, the aggregation at the learning unit level uses the well-known method of mapping the values of an ordered categorical variable to ordinal numbers, as for example explained in (Han, Kamber & Pei 2012) p. 74. The values grey, red, yellow and green in this order are mapped to 0, 1, 2 and 3 respectively. Take the example of a unit with three exercises. Consider a student who solved correctly two exercises – green color code – and did not attempt the third exercise – grey color code. The aggregation value at the unit level for this student is (3+3+0)/3 = 2, which is color-coded in yellow. If the third exercise was wrongly solved, the aggregation value will be (3+3+1)/3 = 2.3, which is also color-coded in yellow. However, if the third exercise was partially correct, the aggregation value will be (3+3+2)/3 = 2.6 and color-coded in green. The same procedure is used to produce the visualizations at the section or course level. This procedure can be applied whatever metrics have been used to produce the visualization of all the exercises of a learning unit. The same kind of visualization has been implemented for the self-assessments on the learning objectives. Teachers can see at a glance how their students assess their own knowledge on each learning objective of a learning unit, and, as above, obtain an overview at the course, section and learning unit level.

For instructional designers as well as for the recommender service, it is important to know whether the typical learning time indicated in the metadata for each learning object is realistic. To this end, we propose a visualization that shows not only the central tendency but also the dispersion of the overall time students spend on a learning object. The visualization is a sequence of simplified box-plots; each box represents a session with the bottom of the box being the minimum time spent by a student on that object in that session and the top of the box the maximal time; average time is drawn as a line in the box; the typical learning time as given in the meta-data is also represented, see Figure 4. On higher levels, such as a learning unit, each box represents the time spent on all learning objects and each student is considered by the overall time spent (sum of all sessions).

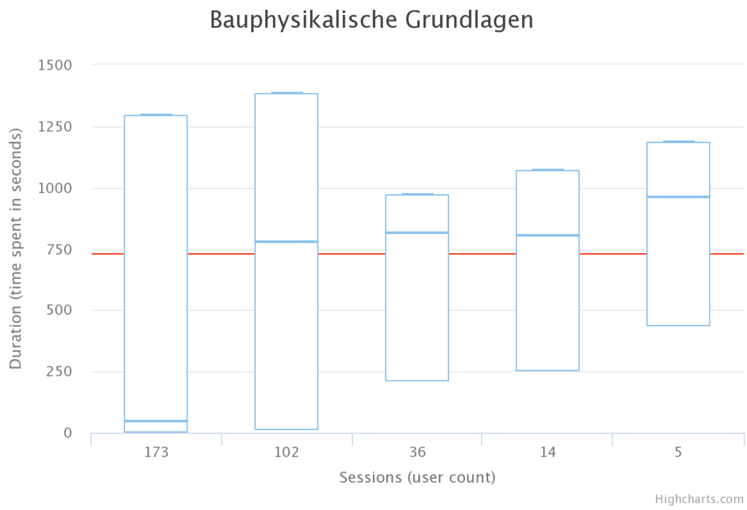


Figure 4: In the first session, 173 students accessed the object. The minimal time spent was about 1 second, the maximum above 1250 seconds and the average about 20 seconds. In a second session, 102 students accessed that object. The typical learning time is given by the red line

4 CONCLUSION AND FUTURE WORKS

The visualizations presented in this paper are for teachers and instructional designers. By showing them how many students are in the green, yellow, red and grey areas, teachers can prepare their next class according to the needs of their students. By showing them evidence of how long students spend on learning resources or how many attempts they make on exercises, instructional designers can reflect on the length and difficulty level of the resources and adapt them. Students can see the same diagrams as teachers do, but with their own data instead of the full class.

An obvious prerequisite for the diagrams to be useful is that the exercises have a high quality and learning objectives are well formulated. This requires some effort. The diagrams support a more active pedagogy style like the inverted classroom. Experience shows that it requires some training for teachers to integrate the diagrams in their daily routine. Further works include diagrams showing to instructional designers the paths that students follow while navigating through the learning objects and the learning units. This learning analytics service is being realized as an LTI-Tool so that it can be used with any LTI-compliant learning system.

Acknowledgements: The authors thank the whole Smart Learning team. This work is partially supported by the German Federal Ministry of Education and Research grant number 01PD17002B.

REFERENCES

Dollár, A., Steif, P. (2012). *Web-based statics course with learning dashboard for instructors*, Proceedings of Computers and Advanced Technology in Education (CATE 2012), Napoli

Han, H., Kamber, M., Pei, J. (2012): *Data Mining Concepts and Techniques* (3rd. Ed.). Elsevier.

Krauss, C., Merceron, A., An, T.-S., Zwicklbauer, M., Steglich, S., Arbanowski, S. (2017). *Teaching Advanced Web Technologies with a Mobile Learning Companion Application.* Proceedings of the 16th ACM World Conference on Mobile and Contextual Learning (mlearn 2017), ACM, Ont. 30 - Nov. 1, 2017, Larnaca, Cyprus, doi>10.1145/3136907.3136937

Visualizing Cronbach's Alpha for a Large Number of Assessments

Aditya Sharma
Playpower Labs
aditya.sharma@playpowerlabs.com

Nirmal Patel
Playpower Labs
nirmal@playpowerlabs.com

ABSTRACT: In this paper, we present a novel data visualization that shows the distribution of Cronbach's Alpha for a large number of assessments and their items. Cronbach's alpha measure can be used to improve assessments by removing items that are not consistent with other items of the test. The exclusion of items can affect the alpha of the assessment in a different way. The proposed visualization makes it easy to identify assessments where the removal of some items can lead to a significant gain in the internal consistency of the assessment items. The visualization is particularly useful when the number of assessments being analyzed is large. The visualization presented is also open source and reusable.

Keywords: Interactive Data Visualization, Assessment Data, Cronbach's Alpha

1 INTRODUCTION

Quality assessments are central to the design of a good curriculum. They provide us with measurements that help us gauge how well students understand the instruction of the course. Assessments become even more important in the context of online curricula, because the instructor is physically separated from her students, and assessment data is one of the few ways to gain an insight into the progress of the cohort. Online learning platforms are now able to employ various kinds of assessments, ranging from in-video quizzes to auto-graded programming assignments to know how well students are moving towards the goals of the course. But the measures of student knowledge provided by these quizzes and tests are directly related to the quality of the assessments. Good quality assessments can tell us right things about students' progress, but bad quality assessments can give us ambiguous information about student knowledge.

There are many different metrics to assess the quality of the tests. Some of these metrics give us information about the quality of test items, while other metrics give us information about the quality of the overall test. Moreover, different test theories such as classical test theory and item response theory give us different metrics to measure how effective the items and tests are to measure student knowledge. The typical measures for items from classical test theory are percent correct and point biserial (or item-total correlation,) while item response theory has its own procedures to calculate item difficulty and discrimination. For tests, classical test theory provides Cronbach's alpha (Cronbach, 1951) that measures the internal consistency of the items, while item response theory uses a measure called test information function that relates test information with latent student ability. All of these measures give us actionable information about how we can improve the test so as to improve our estimate of student ability.

2 METHOD

In this paper, we focus on Cronbach's alpha, which is one of the most widely used metrics to measure the quality of tests. Cronbach's alpha tells how closely items within a test are related to each other. It is a measure of internal consistency. If Cronbach's alpha is very high for a test, we can assume that all of the items in the test are measuring a similar knowledge construct. If Cronbach's alpha is very low for a test, we can assume that the questions of an assessment are trying to measure very different or unrelated constructs. Ideally, for any curriculum, we would want the alpha to be high for every assessment. But often times, we can run into assessments with low Cronbach's alpha. One reason for low alpha is that all of the items are measuring different knowledge constructs, but another reason for low alpha is that only one or some items of the test are outliers, while other items of the test are correlated. If we can remove these outlier items, we can make the test more consistent.

One way to identify items that are not consistent with other items of the test is to look at how Cronbach's Alpha is affected when an item is dropped out of the test. If the alpha of the test increases after removal of an item, we can surmise that the item that was removed helped to increase the internal consistency of the test items. A test with high internal consistency can provide us with a more reliable measure of student knowledge. Imagine a test about for loops in programming that contains some items on if/else statements that students haven't learned about. Results of this test are more reliable if we remove all of the items testing if/else statements, because then, the scores will tell us more about the student knowledge of for loops.

To detect outlier items in a test, we can use item response data of students to calculate the alpha for the test, and alpha for the test after removal of each of the test item. The metrics obtained thereafter can help us decide which items from the test can be removed to improve the quality of the test. This gives us a way to use data to improve the design of assessment, which, in turn, might improve the quality of data collected later.

A Cronbach’s alpha value of 0.70 is considered acceptable (Cortina, 1993). For a small number of assessments, we can calculate Cronbach’s alpha and look at a table of values to find out removal of which increases the alpha of the test. But, when the number of assessments from a curriculum is large (e.g. N > 100,) we might find it very difficult to identify assessments that can be improved the most. We can look at sorted tabular data to find outlier items that can lead to the most improvement in alpha, but we would not get much insight into the distribution of the rest of the data. In this case, it is possible to use a data visualization that helps us identify both the outlier items and overall pattern of the assessment quality. This is the motivation behind the visualization described in this paper.

The visualization presented in this paper is also open source[1] and is written in R, and it can take data in a specific format and generate the visual for any set of input data.

3 DATA

To generate the visualization using the provided open source R code, we need item response data of students for different assessments in a single table. Table 1 describes the data format and shows an example table that can be ingested by the R program to generate the visualization.

The Assessment ID, Question ID, and Student ID column can be any possible unique identifiers for assessments, questions within those assessments, and students who attempted those assessments. Assessment IDs will appear on the X-axis of the plot, so it is suggested to use more meaningful names in that column. For every Question ID in the data, a dot will be made in the visualization and hovering over that dot will reveal back the Question ID. So interpretable values in the Question ID can also be very useful. Student ID column can contain either anonymized IDs or real student names, it will not make a difference in any aspect of the visualization.

Table 1: Data format and example data required to generate the visualization

Assessment ID (String)	Question ID (String)	Student ID (String)	Correct (1/0)	Time (YYYY-MM-DD HH:MM:SS)
Test1	Q1	Anon1	1	2018-01-01 10:00:00
Test 1	Q2	Anon1	0	2018-01-01 10:01:00
Test 1	Q3	Anon1	1	2018-01-01 10:02:20
Test 2	Q1	Anon2	1	2018-01-02 11:30:00
Test2	Q2	Anon2	0	2018-01-02 11:30:55
Test3	Q1	Anon3	1	2018-01-04 15:20:12

4 VISUALIZATION

The visualization is generated by using the data described in Table 1. The visualization program uses the function `alpha()` from R package `psych` (Revelle, 2017) to calculate the Cronbach's alpha. The alpha value is calculated for each assessment inclusive of all items, and the alpha value is re-calculated by removing each item of the assessments. So if an assessment has n test items, n + 1 alpha values will be calculated for that assessment. If there are a total of m unique assessment IDs in the data, m x (n + 1) alpha values will be calculated. Once the calculation is finished, all of these values will be visualized together.

An example of the visualization is shown in Figure 1. The X-axis of the plot shows the Assessment ID. The Y-axis of the plot ranges from 0 to 1 and points the Cronbach's alpha for the assessment on the X-axis. In the visualization, every assessment is represented as a vertical line with multiple dots. The dots are of two colors: blue and black. The blue dots show the alpha of assessment without dropping any item, while the different black dots show alpha of the assessment after dropping one of the

[1] http://github.com/nirmalpatel/edviz-2019

items. A red horizontal line at Y = 0.7 is drawn as a reference alpha value that is generally seen as acceptable. The assessments on the X-axis are sorted from the highest to lowest alpha values that were calculated by using all of the assessment items. The visualization is also interactive, and as shown in Figure 2, hovering over any dot will tell us which item the dot refers to, and how the alpha of the assessment is affected by removing the selected item.

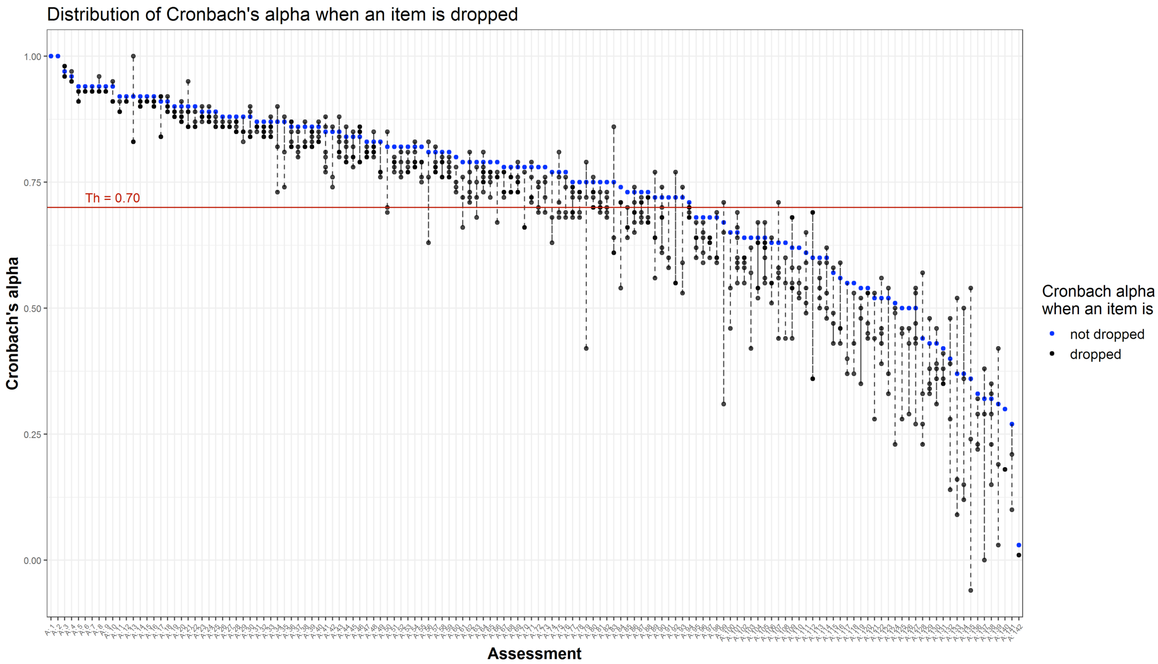


Figure 1: Distribution of Cronbach's alpha when an item within an assessment is dropped. The points in blue color signify Cronbach's alpha for the overall assessment and the points in black signify Cronbach's alpha when one of the assessment items is dropped.

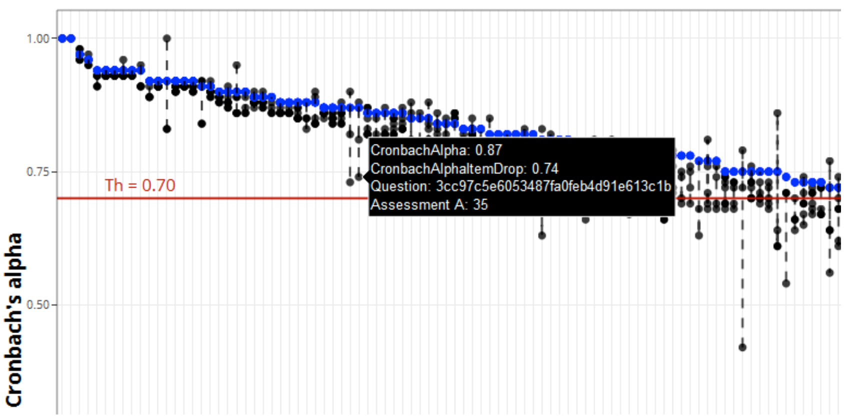


Figure 2: The assessment and item information show up when we hover over the dots of the plot.

5 DISCUSSION

The visualization of Cronbach's alpha for many assessments at once provides us a very easy to grasp overview of the assessment quality data. Rather than just seeing a distribution (or histogram) of Cronbach's alpha for a set of assessments, the visualization enables us to see how different items are impacting the overall alphas of the tests.

The first feature to notice is where the blue dot or the overall alpha of the assessment lies. If this is below 0.7 or another expert defined threshold, the test may be of concern. Another noticeable feature of the visualization is how close all of the black dots are for an assessment. If the dots are very close together, this that the removal of a single item may not affect the alpha of the test significantly. If the black dots of multiple items coincide, it means that the coinciding items contribute similarly to the scale in terms of consistency.

The next important feature of the plot is assessments with wide variation in their alpha values when items are removed. Some black dots in the plot goes high above the blue dots, while some black dots go well down below. The black dots going above the blue dot tell us that removal of an item can increase the alpha of the assessment, while the black dots going below the blue dot tell us that if we remove some items of the assessment, the alpha value can decrease.

In Figure 1, we can see that on the top left the 13th assessment from the left has one item that can be removed to increase the overall alpha, while removal of another item can decrease the alpha. There are a few more assessments between an alpha value of 0.5 and 0.7 that can be improved by removing certain items.

6 CONCLUSION

This described visualization can be used by instructional designers to identify assessments that need attention. Quality tests are a very important tool to gauge student understanding of the learning material, and having a test with high consistency can provide instructors with more reliable estimates of student mastery. This reliability can lead to a better ability to use data to drive instruction.

7 FUTURE WORK

The visualization is still limited in its ability to show other related item analysis metrics such as percent correct and point biserial. These can provide users with more information about how well items in the plot are performing. Another feature that is missing in this visualization is showing how the removal of multiple items can affect the overall alpha. Although this is more complex as there are combinatorial possibilities when removing a combination of items. But adding these features to the plot might help assessment designers track down problematic items in a more better way.

REFERENCES

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Revelle, W. (2017) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.7.8.