On Instructional Utility, Statistical Methodology, and the Added Value of ECD: Lessons Learned from the Special Issue

BRIAN NELSON
Arizona State University

REBECCA NUGENT
Carnegie Mellon University

and

ANDRÉ A. RUPP
University of Maryland

This special issue of JEDM was dedicated to bridging work done in the disciplines of *educational and psychological assessment* and *educational data mining* (EDM) via the assessment design and implementation framework of *evidence-centered design* (ECD). It consisted of a series of five papers: one conceptual paper on ECD, three applied case studies that use ECD and EDM tools, and one simulation study that relies on ECD for its design and EDM for its implementation. In this reflection piece we discuss some of the key lessons that we have learned from the articles in this special issue with respect to the instructional utility of the digital learning environments, the nature of the statistical methodologies used, and the added value of the ECD framework for the work conducted in these projects.

Keywords: Educational data mining, psychometrics, digital learning environments, statistical methods, instructional utility, special issue

---

Authors' addresses: Brian Nelson, School for Computing, Informatics, and Decision Systems Engineering, Arizona State University, Brickyard Engineering M1-04, Tempe, AZ, 85287, brain.nelson@asu.edu; Rebecca Nugent, Department of Statistics, Baker Hall 132, Carnegie Mellon University, Pittsburgh, PA, 15213, rnugent@stat.cmu.edu; André A. Rupp, Department of Human Development and Quantitative Methodology, University of Maryland, 1230-A Benjamin Building, College Park, 20742, ruppandr@umd.edu

1. DIGITAL LEARNING ENVIRONMENTS FOR LEARNING AND ASSESSMENT

As an instructional designer focused on the use of immersive virtual environments for learning and assessment, it is very exciting to be able to review some of the excellent work taking place around the ECD framework and EDM used with digital learning environments. There has been much enthusiasm recently around the role of “big data” and digital learning environments. As the articles in this special issue clearly demonstrate, strong theory-based application of the ECD framework combined with sophisticated EDM techniques and psychometric analysis can provide a reliable, valid, and scalable approach to better understanding learning trajectories at the individual and aggregate level.

This work comes at an interesting time for assessment policy in the United States. There is growing dissatisfaction among researchers and policy makers around the issue of existing standardized assessments for K-12 students. Traditional assessments are increasingly used as tools not just for evaluating student mastery of content, concepts, and processes, but also as evaluative measures of the success of schools to teach their students. Consequently, it is more important than ever to ensure that assessments and associated analysis methods are valid and meaningful. With traditional forms of assessments with highly constrained task types (e.g., multiple choice, true-false, fill-in-the-blank, or short answer questions), students are evaluated based on analysis of a remarkably small amount of data. Statistical methodologists have nevertheless made use of sophisticated analytical tools to glean insights about student learning from a small trickle of available data.

With the advent of more sophisticated digital learning environments, the situation has been utterly changed. Particularly with the kinds of immersive environments and simulations described in this issue, the empirical faucet has opened wide resulting in an overwhelming torrent of data. Every interaction including mouse clicks, individual keystrokes, text passages, voice-based data, virtual location information, and more can now potentially be recorded and analyzed, both in the moment and over time. Analysis and visualization of this data can provide a steady stream of information to students and teachers throughout the learning process, giving both groups better insights into student understanding.

Faced with this wealth (or overabundance) of data, the researchers featured in this special issue have found that ECD provides a useful framework for the creation of activities and work products in digital learning environments, and as a guide for making sense of the data emerging from these activities. Further, as the case study articles in this issue show, the combination of the ECD framework with EDM techniques can offer further power for an iterative cycle of design, analysis, and redesign of digital learning environments as platforms for assessment.

As is appropriate to the theme of this special issue, the studies presented focus mainly on the application of ECD and EDM methods to support the analysis of existing data derived from digital learning environment-based curricula: in other words, data output. However, as discussed most directly by

Mislevy et al. in their lead article, it is also important to consider carefully the role of the ECD framework in structuring the activities, tasks, and work products that are designed into digital learning environments. These are designed spaces, with specific theory-based ideas guiding their creation. Consequently, researchers have the opportunity to "shape the stream": to design learning activities, tasks, and the learning environment itself in a way that reduces noise in the outcoming flow of data.

On a p ractical basis, shaping activities is fairly straightforward. Despite their seeming complexity, digital learning environments have a finite number of possible channels for data input. For example, digital objects and artifacts embedded in learning environments provide a main source of data for analysis of learning. By careful application of the ECD framework in the initial design and presentation of such objects in a given learning environment, the task of analyzing outcoming data becomes more straightforward. With foreknowledge of the palette of possible objects, and hypothesized meanings associated with individual and aggregate interaction with available objects, researchers can more easily interpret such things as the order or timing of student interactions with digital objects, the choice of objects used to complete specific tasks, or the selection and assembly of simple objects into more complex ones.

2. THE NATURE OF THE STATISTICAL METHODOLOGIES

As mentioned above, one of the fantastic, but possibly frustrating, by-products of the richness of today's digital learning environments is the vast quantity of data steadily being generated and stored. We say *fantastic* due to the unparalleled opportunities to model complex learning processes and optimize assessment procedures. However, an overabundance of possibly unnecessary data and information can *frustrate* even the most thorough instructional designer in search of a signal in the presence of so much potential noise. Given the decreasing costs of data collection and storage, we need to start asking ourselves if this overabundance of data is necessarily a good thing. We may have a digital ocean, but what if we have been keeping track of all the wrong things? While simply storing anything and everything and sorting it out later seems like an attractive solution (following the philosophy of better to have it and not need it than to need it and not have it), we may quickly finding ourselves drowning in data while searching for the proverbial needle in the haystack.

The case studies presented in this special issue are illustrative examples of utilizing and incorporating modern statistical methodology designed to identify (possibly weak) signals in the presence of noise. In particular, EDM or machine learning tools are frequently being used to analyze these learning environments, including modeling different student cognitive learning processes, identifying possible group structure in measures like student performance, problem difficulty, classes, assigned tasks, and so on as well as determining which metrics to track and when to track them. While more traditional psychometric approaches may offer more theoretically consistent results, they are much less adept at handling the high dimensionality of today's problems.

In addition, given the need for immediate (or near immediate) feedback and assessment in adaptive digital learning environments, any incorporated statistical methodology needs to provide real-time results that can be easily summarized for use by both the system designers and users. This is not an easy task, but one on which the included case studies are showing some real progress. To move forward, we believe that the field will need to embrace the strengths of these high-dimensional approaches while seeking the desired theoretical properties of the more traditional psychometric models. Otherwise, we will be forced to choose between what some view as "approximate solutions" and only having the ability to analyze "small problems". In all cases, summarizing and presenting our results cannot be an afterthought. An open dialogue in the community is needed.

The studies in this special issue also all recognize the need to incorporate advanced statistical methodology early and often in the design and analysis of their respective digital learning environments. Each project has been designed and implemented by an interdisciplinary team that includes members with extensive statistical expertise. In some cases, statistical experiments/studies are being performed in early design stages to better inform the overall structure of the digital learning environments. This type of iterative process is indicative of the adaptability that will be necessary to have digital learning environments that evolve as needed given a new set of goals or possibly a new framework. In addition, without having appropriate statistical training or an interdisciplinary team with the right background, it is too easy for analyses to turn into "just blindly try everything and see what features come up as significant", a poorly motivated statistical philosophy that can lead to inconsistent results and misleading conclusions. The teams here are to be commended for showing foresight in their statistical planning; early studies can result in more efficient learning environments that focus on reliably answering scientific questions proposed in advance.

As a whole, the included articles indicate the large potential in combining appropriate statistical methodology and educational data mining tools with today's digital learning environments. Although a project's primary analysis goal might be, for example, diagnostic assessment in the classroom, methodologists should not pass up the opportunity to develop new statistical models and approaches within this joint ECD/EDM framework. Given the wealth of available data, the richness of the learning environments, and the interdisciplinary background of the researchers, it seems entirely plausible that this field, rather than being confined to "just data analysis", could produce the next big methodological breakthroughs; we certainly encourage everyone to try!

3. THE ECD FRAMEWORK IN PRACTICE

All of the articles in this special issue have demonstrated some type of added value of the ECD framework for designing and implementing diagnostic assessment features for digital learning environments. As we noted in our introduction to this special issue, the field of educational and psychological assessment in particular has seen a steady increase in the popularity of the ECD framework. This has led researchers and practitioners on a path toward making new meaning of the somewhat abstract descriptions of the ECD components that are presented in the more theoretical literature on ECD. With this special issue we have put into practice our belief that the best way of understanding the power of the ECD framework is through annotated case studies.

Despite the fact that the ECD framework provides interdisciplinary teams of experts with a consistent language for discussing and operationalizing their choices for assessment design, presentation, and data analysis, there is still some ambiguity left with respect to how the different layers, models, and subcomponents should be enacted in practice. Consequently, during the creation of this special issue, the authors of the papers and we, as editors, were sometimes renegotiating the usage of specific terms across papers in order to create coherence for the alignment of ECD concepts and practical instantiations of them. Based on this experience and numerous discussions with colleagues, we expect that it will still take a while before more theoretically-oriented researchers and more practically-oriented specialists agree on a singular vocabulary for the ECD framework.

The three case studies and the simulation study in this paper have also helped to illustrate how different ECD components get enacted in practice. While it is relatively easy to read a theoretical description of the ECD framework and agree with the presented rationale, it may not be clear to researchers who design digital learning environments what representational forms objects in the different layers of ECD take. The papers in this special issue have either explicitly noted or alluded to the fact that these objects include text documents, spreadsheets, presentations, papers and technical reports, Java applets, dynamic databases, modules of specialized code, and so on. Furthermore, even for projects who work with a similar type of work product from learners such as log files, the actual ways these get stored, transformed, and analyzed as part of the same ECD model is different across projects. The power of the ECD framework is to allow for such surface-level differences and to help understand the deep-structure differences across projects vis-à-vis the evidentiary reasoning process. Over time, using an increasing number of annotated case studies, it is likely that we arrive at certain prototype applications of the ECD framework and some that bear an accepted family resemblance to these.

Even with some remaining semantic ambiguity surrounding some of the ECD terminology and the way components get enacted, the papers in this special issue have demonstrated that this framework can help members of the interdisciplinary terms reason more sharply through an evidentiary assessment argument. Put differently, even if members of an interdisciplinary team tend to use slightly different terms and approaches for operationalizing key design features, the meaning-oriented re-negotiation process anchored by the ECD framework that is necessary to make projects successful in practice is a useful professional development exercise in and of itself. This is clearly critical for design-based projects with digital learning environments where members with different types and levels of training have to collaborate constructively. Consider, for example, that none of the papers in this special issue were single-author publications and that almost all cited work in this area within the papers reflects the work of teams as well.

Another feature of the ECD framework that has often been highlighted is its relatively value-neutral stance toward the kinds of assessments that are being designed and the theoretical frameworks that guide their development and score interpretation. For example, the ECD framework is equally suitable for reasoning through the design, implementation, and data analysis for those digital learning environments that focus on socio-cognitive narratives about groups of learners as it is for those digital learning environments that focus on information-processing narratives about individual learners. While the applications in this special issue were predominantly concerned with individual learners, it would be important to showcase more instantiations of the ECD framework for learning and assessment activities that are used to characterize groups of learners.

The ECD framework can also help to reconcile perspectives from different methodological / statistical disciplines. As a tool that supports the creation of a coherent validation argument for a specified purpose the ECD framework embraces qualitative, quantitative, and mixed-methods approaches for gathering evidence about individual learners and / or groups of learners. For example, qualitatively-driven inquiry tools typically require very different methodological approaches for data collection and analysis than quantitatively-driven approaches. Yet it is still important for interdisciplinary team members to agree on how one would like to characterize the learners (i.e. the student model), what information the tasks can provide for such characterizations (i.e. the task model), and how relevant information from the data about learner characteristics should be identified and compiled (i.e. the evidence model).

Similarly, as discussed by Mislevy et al. in the lead article in this issue, there is a fertile inter-methodological space where statistical tools from multiple disciplines, including EDM, can productively co-exist. Moreover, as we argued in our own introduction to this issue, the disciplinary boundaries for different statistical methods are often more fuzzy than concrete with a lot of productive theoretical and practical carry-over. The ECD framework supports an assessment engineering perspective from which the disciplinary foundations of particular methods matter less than their productive use in the service of creating a coherent evidence-based narrative of learners. The quest for deep-structure commonalities across perspectives and tools that the ECD framework supports may be discomforting to those who feel it is most important to defend ideological stakes within seemingly well-defined disciplinary boundaries. We would argue, however, that this is less productive for interdisciplinary research on digital learning environments than leveraging the unique perspectives of the work in different disciplines for a common assessment goal. All three case studies in this paper as well as the simulation study have demonstrated how powerful the ECD framework can be for accomplishing these goals.

We want to close this section of our reflection piece by reminding readers that the ECD framework is certainly not the only mechanism that could be used to achieve the above goals; however, it seems like a rather attractive one to us. It would be interesting to hear from readers to hear about points of disagreement and contention to see for what contexts certain features of alternative frameworks might be more suitable for accomplishing similar development and implementation goals. We thus hope that this special issue will become part of a broader endeavor for showcasing the power or principled assessment design and implementation frameworks like ECD for the development of embedded assessments for digital learning environments. We are looking forward to hearing from our readers about their experiences, reflections, and suggestions for future work!

Sincerely,

Brian Nelson, Rebecca Nugent, and André A. Rupp
(Co-editors, JEDM Special Issue)