Computational Psychometrics Approach to Holistic Learning and Assessment Systems

Alina A. von Davier [1], **Benjamin Deonovic** [1,*], **Michael Yudelson** [1] **Stephen T. Polyak** [1], **and Ada Woo** [1]

[1] *ACTNext, ACT, Inc., Iowa City, IA, USA*

Correspondence*:
Benjamin Deonovic
benjamin.deonovic@act.org

ABSTRACT

Learning and assessment systems have grown and taken shape to incorporate concepts from both models for assessment and models for learning. In this paper we argue that a third dimension is necessary. Not only is it important to understand what the capabilities of a learner are, and how to grow and expand these capabilities, but we must consider where the learner is headed; we need to consider models for navigation. This holistic perspective of learning and assessment systems is encapsulated in the extended learning and assessment system, a framework for conducting research. Fundamental to this framework is the role of computational psychometrics to facilitate the abstraction from raw data to conceptual models. We provide several examples of research projects and describe how they fit into the described framework.

Introduction Many of the characteristics of today's classrooms would be familiar to our great-grandparents: A teacher lecturing to students sitting in rows of organized desks; The teacher instructing from a prepared lesson plan, and the students listening attentively. This traditional education system has been in place with almost no perceptible change since the dawn of the previous century. Students are grouped into various hierarchical aggregations such as classrooms, grades, and schools. The education that students receive is then primarily tailored to these groups as a one size fits all approach rather than a personalized and adaptive experience. The fiction author William Gibson said, "The future is already here, it is just not very evenly distributed" (Rosenberg 1992) In his quote Gibson alludes primarily to the fact that progress is simply the spread of what is niche to something that is ubiquitous and equitable. This could be said of the state of learning and assessment systems in the current era. Recent advances in computing technology have given us the tools to realize many innovative ideas previously beyond our grasp. Many of these innovative ideas are in the various fields associated with education, learning, and assessment. The new discipline Computational Psychometrics (A A von Davier 2015; A A von Davier 2017) sits squarely in the intersection of these fields. Computational psychometrics describes the blend of the analytical tools from the machine learning (ML) arsenal with cutting edge work in theoretical psychometric research. Advances in ML and big data analyses have allowed psychometric researchers to incorporate these tools to form the computational psychometrics paradigm. Computational psychometrics is currently being applied to a range of learning and assessment research topics, from collaborative problem solving skills (S T Polyak et al. 2017) to

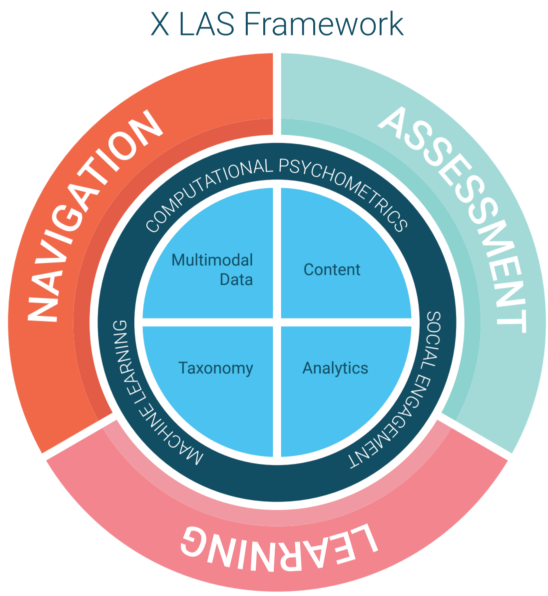


Figure 1. Extended Learning and Assessment System (XLAS) Framework

the impact of interpersonal communications on reciprocity in economic decision making (Cipresso et al. 2015) and to learning as we describe here. Computational psychometrics explores not only novel models for new data types, such as complex process data, but also how these models can integrate and make components of teaching, learning, and assessment more holistic and connected. In this paper we highlight the need for a framework for educational applications and practices which takes a holistic approach to assessment. Such a holistic approach needs to not only blend together models for assessment with models for learning, as has been previously suggested (See for example Tomlinson (2004)), but also include models for navigation, which, broadly speaking, refer to a holistic approach to manage the educational options available to the learner. We begin by describing the eXtended Learning and Assessment System (XLAS) which is our proposed framework for blending these three disciplines, assessment, learning, and navigation, in a holistic manner. Within the XLAS framework computational psychometrics provides 1) theoretical and practical foundations (e.g., learning theory, measurement rubrics, developmental trajectories, etc.) for the ongoing development of the framework and 2) computational and analytic tools for using evidence collected from the applications of the framework, to validate and improve the framework and its underlying theory, curriculum, and algorithms. Next we proceed to enumerate several concrete and ongoing projects that live in the XLAS space. Each project utilizes various aspects of computational psychometrics to bring together independent systems and blend two or more of learning, assessment, and navigation. We will highlight specific aspects of these projects that utilize computational psychometrics. We conclude with a brief summary along with a vision for the future of educational research.

Extended Learning and Assessment Systems

We are all experienced learners. From our earliest experiences grasping the concepts of movement and basic language acquisition, to our forays into arithmetic, grammar, and later, more complex constructs such as time management and teamwork, we spend a significant portion of our lives learning the skills necessary to navigate our world. The systems and frameworks that encompass and define how our learning occurs are called learning systems. Learning systems take on a variety of forms from traditional examples such as classrooms, textbooks, and apprenticeships, to more modern adaptations such as computers and online forums.

Learning and assessment are intricately linked in a person's journey of acquiring new skills or expanding one's abilities. While learning is the process by which a person gains knowledge or skills, assessment is a way to observe the performance in a learner and produce data in such a way that inferences may be made about what the learner has learned. A particularly succinct description of the relationship between learning and assessment is that effective assessment supports learning by providing evidence (1) of learners achieving learning goals, (2) to inform teachers' decisions, and (3) that reflects teaching effectiveness and informs future instructions (Suskie 2018) The relationship between the learning system and the assessment system may range from being completely independent to intimately related and tied together in a feedback loop in which one system provides information to the other. We call this joint system of learning and assessment the Learning and Assessment System (LAS).

In the current paper, the authors propose that there is one additional critical component in the learning and assessment loop: Navigation. Navigation is the ability to find a path from one's current state to a goal state. Navigation includes social emotional learning (SEL) skills and decision making skills, which together with the academic skills, support one's success in education and workforce in a holistic manner. We refer to the system which includes the components of learning, assessment, and navigation as the eXtended Learning and Assessment System (XLAS). In the context of the XLAS, navigation refers to the ability of the learner to utilize the affordances available from the system to make the right choices during the process of learning. Examples of education and career navigation skills may include time management skills, self-knowledge of abilities and interests, knowledge about academic major and occupations, and skills related to planning and decision-making (Camara et al. 2015) Navigation components of an XLAS could include teachers as the curators of knowledge, virtual agents, or system-based affordances, such as recommendations and learning analytics. This navigation component may interact with both the learning and assessment subsystems by curating or designing learning experiences or designing and administering an assessment. We have worked to develop models and systems that integrate these three components of the XLAS in a holistic manner. Figure 1 illustrates the interactions among the XLAS subsystems and possible components in an XLAS. Learning, assessment, and navigation all interconnect. Each subsystem interacts and informs the other two. The center of the graph portrays lower level features and direct derivatives of these lower level features including multimodal data, content, taxonomies, and analytics (See Section 0.1). The outer ring of Figure 1 portrays the three main subsystems of the XLAS: learning, assessment, and navigation. These subsystems correspond to higher level abstract models. Examples include Item Response Theory (IRT) for assessment (van der Linden 2018) Knowledge Space Theory (KST) for learning (Doignon and Falmagne 1999) Holland's Theory of Career Choice for navigation

(Holland 1958) The inner ring represents the paradigms that allow researchers to link the lower level features and derivatives to higher level abstract models Khan (2017) These paradigms include computational psychometrics, machine learning, and social engagement. A A von Davier (2017) argues that the main feature of computational psychometrics is that the data collection is intentional and by design, hence theory-based. In this way computational psychometrics allows researchers to form links between the higher level abstract models to the concrete components at the center of the XLAS in a top-down manner. The machine learning paradigm on the other hand allows one to abstract the concrete components in a bottom-up manner by utilizing algorithms to build predictive models. Additionally, social engagement in learning, assessment and navigation refers to the degree that an individual participates in these systems within a particular community or society. Such participation can further help define and refine the links and connections between lower level features and higher level abstract models. An example of research, in collaborative problem solving, which utilizes social engagement to help define the links between the data and the abstract model is presented in Stoeffler et al. (2017)

In the current paper we will illustrate each of the subsystems of the XLAS framework by using current literature and research projects as examples. In each example we highlight how we connect lower level features to high level abstract models. We will also further explore specific building blocks of the XLAS and summarize use cases that address these intersections among the XLAS components.

Building Blocks and Use Cases

0.1 Data, Taxonomies, Content, and Analytics

Here we delve deeper in explaining the XLAS from 1. The subsystems of learning, assessment, and navigation in the XLAS represent high level, complex constructs and models. In the center of the framework are lower level features and derivatives of lower level features. Lower level features include multimodal data and metadata. Examples of multimodal data are audio, video, and sensor-based data as well as more traditional assessment data such as response data. Metadata includes additional covariate information; for response data it may include to what particular taxonomic learning standards an item has been tagged to or the demographic information for the learner. Derivatives of lower level features include content, taxonomies, and analytics. Lower level features are connected to higher level complex constructs through a series of hierarchical abstractions. From the data, content, and taxonomies particular relevant features are extracted. These features are combined into mid-level representations. These mid level representations are then used directly in the models of complex abilities such as learning or navigation. Computational psychometrics through psychometric models, machine learning, and social engagement solutions, serves as the paradigm which connects these lower level features to higher level features. One focus of the research on the XLAS framework revolves around identifying, constructing, or obtaining useful low level features from the multimodal data and content; another focus is on feedback and analytics. In the following sections we provide two examples of research associated with these aspect of the XLAS.

0.1.1 Theory-Based taxonomies and standards

The most important difference between computational psychometrics and machine learning is that the data collection in computational psychometrics is intentional and aligned with a theoretical framework or taxonomy(A A von Davier 2017) A taxonomy specifies how specific knowledge, skills, abilities and other characteristics connect to broader domains of learning. Such a taxonomy is a key abstraction which allows low level features such as response data to be connected to higher level constructs such as learning and educational success. A recent example of the development of such a taxonomy is the Holistic Model of Education and Workplace Success, also known as the Holistic Framework (HF) by Camara et al. (2015)

Policymakers and accountability systems have for a long time focused on academic measures when discussing college and career readiness. However, it is becoming increasingly clear that performance in college and in the workplace depends not only on the traditional academic measures, but also on other socio-emotional and behavioral skills. Camara et al. (2015) identified the need for a framework which gives structure and organization to the knowledge and skills necessary to succeed. Based on a comprehensive review of relevant theory, education and work standards, empirical research, input from experts in the field, and a variety of other sources, they have developed the HF, a comprehensive framework that states what people need to know and be able to do to be successful throughout the course of their education and careers.

The framework is organized into four broad domains: core academic skills, cross-cutting capabilities, behavioral skills, and education and career navigation skills. One of the major facets of the HF is the core academic skills. This section of the HF defines the hierarchical relationship of skills that learners are expected to learn during high school in the domains of language arts learning, mathematics, and science. This provides a core lower level feature in the XLAS framework, namely a taxonomy of knowledge, skills, and abilities, which allows researchers to build complex models of learning, assessment, and navigation. Another key aspect is the developmental nature of the HF. This is important because the precursors of success emerge very early in life and development continues well beyond the confines of traditional LAS.

The development of the HF was fundamental to help bootstrap research in the XLAS framework. For example, developing a set of content related resources for learning that allows for best practices should rely on the HF to define the goals for learning, the knowledge structure, and scaffolds that should guide the students through the learning process. With the advent of such a taxonomy, researchers are able to connect response data from students to particular models of learning, assessment, and navigation. See Section 0.5 for an example of such research which harnesses the HF.

0.1.2 A data cube for educational data

In the previous section we saw how computational psychometrics built upon the development of a taxonomy of hierarchical standards which is a key fundamental lower level feature that helps bootstrap model development in the XLAS framework. Further research has utilized computational psychometrics to refine the analysis of multimodal data to extend the models used in traditional pscyhometrics to identify lower level features that are primed for use in the XLAS framework. This research was sparked by the fact that, in recent years, the work with educational testing

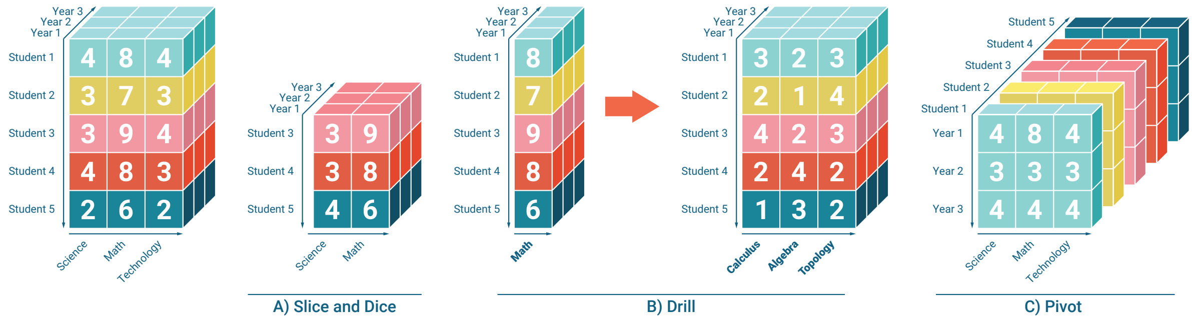


Figure 2. Data Cube for Educational Data. Adapted from A A von Davier et al. (2019) and used with permission under the Creative Commons Attribution (CC-BY) license.

data has changed due to the affordances provided by the technology, the availability of large data sets, and by the advances made in data mining and machine learning. The computational psychometrics paradigm allows researchers to create new connections between theoretical models and these new data features. However, the way data (from multiple tests at multiple times) has been collected, stored, and analyzed by testing organizations is not conducive to these real-time, data intensive computational psychometric and analytics methods that can reveal new patterns and information about the student. In A A von Davier et al. (2019) the authors propose a new way to label, collect, and store data from large scale educational learning and assessment systems (LAS) using the concept of the 'data cube' to relate and align multiple databases. The 'data cube' idea has evolved over time, but the paradigm remains easy to grasp.

One of the ideas proposed by A A von Davier et al. (2019) is to rewrite the taxonomies and standards as mathematical vectors, and add these vectors as dimensions to the 'data cube.' Similarly, they recommend to vectorize the item's metadata and align them on different dimensions of the 'cube.' Psychometricians and data scientists can interactively navigate their data and visualize the results through slicing, dicing, drilling, rolling, and pivoting. A simple example of these various operations can be seen in Figure 2. The drilling-down operation illustrates that the data cube is not necessarily just a multidimensional vector. It can be seen in Figure 2 that dimensions in the data cube can also hold other metadata regarding that dimension. For example the 'subject' dimension Math has associated metadata that corresponds to the hierarchical topics associated with mathematics (in the example this includes Calculus, Algebra, and Topology).

The data cube structure works well with data exchange standards, such as the IMS Global set of standards. In principle, these standards propose data schema for various data features that the users agree upon. For example, the IMS Global Caliper standard is a template for event data collected during the process of a performance task or a learning session.

The data cube and the data standards allow for real-time big data analyses, including the use of ML and computational psychometric techniques for the alignment of testing instruments, real-time updates of cognitive diagnostic models during the learning process, and real-time feedback and routing to appropriate resources for learners and test takers. The fundamental ideas behind the data cube guide many of the authors' current research projects. Specifically in Section 0.5

and Section 0.4 we will see how two projects, the ACT Recommendation and Diagnostic API and the ACTNext Educational Companion App, use the various dimensions of the learning and assessment data to provide learners with deeper insights into their skills and help them navigate educational learning resources.

0.2 Learning and Assessment

Next we will focus on the outer ring of the XLAS framework, which consists of the three subsystems: learning, assessment, and navigation. Rather than describe in detail each subsystem we will describe the possible overlaps between these three subsystems of the framework since the framework allows research to be multifaceted. The first overlap we will discuss is the overlap between learning and assessment. Research that falls under this umbrella focuses on defining the feedback loops between systems of learning and systems of assessment. A large amount of literature is available for each of these systems on their own, but only within the past few decades have researchers started to discuss how the models in learning and the models of assessment can inform each other. Part of this work is theoretical in nature. Note that theoretical models for learning and models for assessment data have diverged and grown to leverage the salient features and distinct assumptions that embody their respective data sets. Assessment models were built to analyze cross sectional data, whereas learning models were built to analyze longitudinal data. Yet, despite the divergent development of these models there is an intimate connection between the two prominent models in these fields: Bayesian Knowledge Tracing (BKT) and Item response Theory (IRT) (Deonovic et al. 2018)

Other work on this overlapping relationship blends theory with practical need. To further explicate the overlap between systems of assessment and systems of learning we consider an extension of the well known evidence centered design (ECD) framework for designing assessments (Mislevy et al. 1999) This extension, the extended ECD (Arieli-Attali et al. 2019) provides room for designing systems in which learning and assessment co-exist.

0.2.1 Extended Evidence Centered Design

As mentioned before, the most important difference between computational psychometrics and machine learning is that the data collection in computational psychometrics is intentional and aligned with a theoretical framework (A A von Davier 2017) Several theoretical frameworks have been developed for assessments. ECD is one such framework designed to place priority on the collection of validity evidence from the onset of the design of the assessment (Mislevy et al. 1999) The three core components of the ECD framework include the Student Model, Task Model, and Evidence Model. The Student Model specifies the latent competencies that are the target of the test, the Task Model specifies the task features that will elicit the observed data which will allow for inference about the latent competencies, and the Evidence Model makes the connection between the latent competencies specified by the Student Model and the observed data from the Task Model. This framework however does not specify how the learning component of an LAS should be designed and developed in order to properly elicit validity evidence. An extension of the ECD framework, the Extended Evidence Centered Design (e-ECD) framework broadens each of the three core components of the ECD as well as draws upon data driven techniques and computational psychometrics to power these extensions (Arieli-Attali et al. 2019)

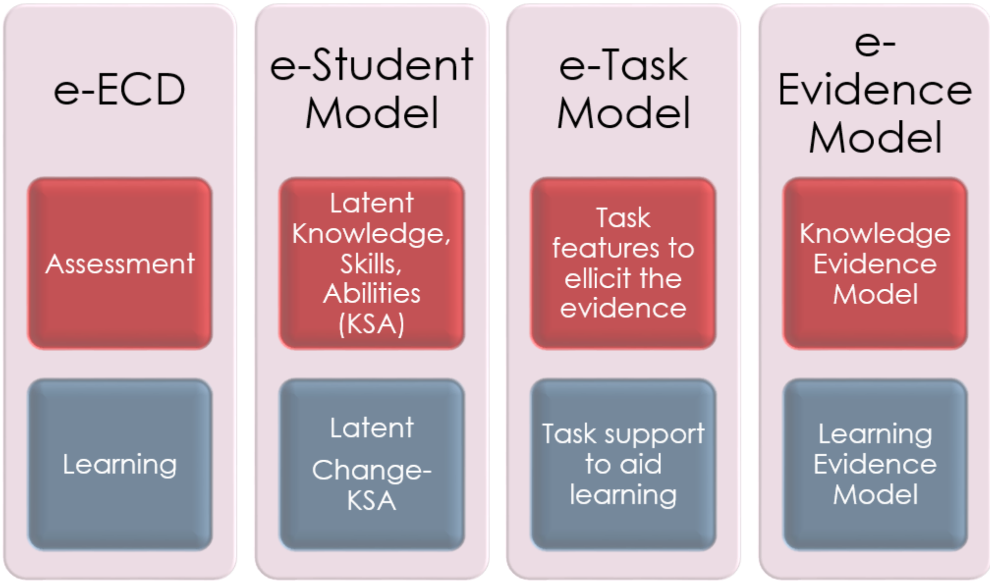


Figure 3. Extended Evidence Centered Design (e-ECD). Adapted from Arieli-Attali et al. (2019) and used with permission under the Creative Commons Attribution (CC-BY) license.

The extended model includes a static layer, corresponding to the original components of the ECD, and a dynamic layer which addresses learning, see Figure 3. The extended components include: (1) the extended Student model, or the Knowledge-Change model, which specifies learning processes as the latent competency that the system is targeting; (2) the extended Task model, or the Task-Supports Model, which specifies principles and features of learning supports (scaffolds, feedback, hints, etc.) that guide the design of tasks; and (3) the extended Evidence model, or the Knowledge-Skills-Abilities Change model, which specifies the links between the students' responses, scaffold usage, and the target learning processes. These links allow for inference from behaviors to latent learning.

Using computational psychometrics and empirical data we can monitor the use and impact of learning supports and dynamic models of ability. This data driven approach combined with the theoretical perspective will help us to create a relevant and well designed framework for the development of learning and assessment systems.

0.3 Navigation and Assessment

Another overlap we are considering is that of navigation and assessment. If navigation can broadly be described as determining where one should go, the subsystem of assessment in this context revolves around knowing were one is at currently. Preparing individuals for the decisions they will be making throughout their educational and career journeys, along with optimizing these decisions, are important areas of study that have the potential for significant impact. We are looking at these aspects from many perspectives that reflect the computational psychometrics paradigm.

We reached out to our colleagues who worked on the navigation part of the ACT Holistic Framework to understand the role of navigation for success. Educational success depends on many factors, including what individuals know about themselves and their environments, and how they use this information to make choices, plan actions, optimize resources, and move along their education and career paths (Becky Bobek, personal communication, December 14, 2018). The literature on decision making processes related to critical navigation decision points such as educational and occupation choices is scant, and we have been working to uncover some of these processes. For example, economists (Wiswall and Zafar 2015) refer to “unobserved tastes” as a dominant factor in the choice of major and reinforce the need to investigate these more heterogeneous aspects of major choice.

Bobek and Moore (2017) refer to the four dimensions critical to navigating education and career transitions effectively encompass

1. Self-knowledge (understanding of one’s abilities, interests, skills, values, attitudes, and beliefs),
2. Environmental Factors (education/work knowledge, experiences, supports, barriers),
3. Integration (exploration, goals, congruence, education/career choice making, action plans),
4. Managing Career & Education Actions (college/job search, roles, implementation).

Research concerning the first three dimensions mentioned above may be found in Paek and Bobek (2018) Separately, research has been conducted on learning analytics, recommendations, notifications, and data display (Whitmer et al. 2017) and on how to model the choices that people make (Kruis 2018) Additional research is being conducted on the optimization of decision making process for individuals and their goals (A von Davier and Arielli-Attali 2018)

0.4 Learning and Navigation

The final overlap on the XLAS framework we will consider is learning and navigation. Navigation refers to knowing were one wants to go and determining how to get there. This is quite abstract as it can apply to many situations. Navigation can refer to deciding on which career to pursue and how to pursue it, deciding which college to go to and what major to enroll in, or even on a more day to day basis navigation can refer to a teacher’s lesson plan, the goals they want their students to achieve and how to achieve them. Learning is the process of going from where one is to where they want to go in terms of the further development of one’s knowledge, skills, and abilities. Research involved in the subsystem of learning in the XLAS deals with constructing theoretical models for how learning occurs, measuring what learning is occurring, and promoting learning.

There has been heavy investment into exploring the intersection between these two subfields in the last few decades with the increase in interest in learning analytics. In education, Aguilar (2018) writes that learning analytics has emerged as the discipline associated with analyzing and reporting big data. Stakeholders in education utilize learning analytics as a way to make learning and learners’ navigation more personalized. Aguilar (2018) points out that most learning and navigational infrastructure is primarily built to serve the "average" student, e.g. students are grouped into classrooms, grades, and schools and teachers provide instruction one class at a time. More personalized instruction has been shown to be impactful, but can be difficult to scale. Aguilar (2018) argues that learning analytics provides a solution to this problem by using computational methodology and visualizations to allow personalized learning and navigation at scale. Methods for learning analytics include:

1. Resource Analytics - resources which students use (from our offerings) and create (such as essays)
2. Behavior Analytics – time on task, persistence, curiosity, participation, etc
3. Social Learning Analytics - social interaction in learning, participation in learning networks, forums, etc
4. Predictive Analytics – within student: timely identification of at-risk behavior; and across students: identify at-risk students in terms of engagement and retention for meaningful intervention before they go off-track

One platform which has successfully utilized learning analytics to bridge together navigation and learning is the ASSISTments project. The ASSISTments project (N T Heffernan and C L Heffernan 2014) provides a system for teachers and researchers to work together. It is a project that is perfectly situated in the learning and navigation overlap. One particular functionality which allows ASSISTments to combine learning and navigation is the ability for teachers to try out various instructional content, learning resources, and interventions on groups of children and monitor and track their learning. This allows researchers and instructors to observe the impact of the content, resources, and interventions on student learning.

However, Aguilar (2018) points out learning analytics is not a silver bullet for the problem of providing personalized learning. The data that are collected and the procedures for processing and creating visualizations needs to be thought through ahead of time. This means that at a particular point in time, the data necessary to truly provide an insightful personalized experience for a student is not available. Furthermore, many learning analytics approaches utilize machine learning techniques such as clustering or visualizations that compare particular students to other aggregates, both wash out individual characteristics and seem counter to the goals of personalized learning analytics. The insights and personalization provided by learning analytics also requires trained experts. For example Teasley (2017) points out that providing learning analytics dashboards directly to students has not been very fruitful and requires further research. Finally, Aguilar (2018) points out potential data privacy concerns, which remain to be comprehensively addressed.

XLAS Use Case: The ACTNext Educational Companion App The XLAS framework successfully links learning, assessment and navigation by managing the relationship of learner's data to content (assessment and instructional), taxonomies (knowledge, skills) and analytics. The authors applied this management strategy in the development of one of our recent research-based prototypes, the ACTNext Educational Companion App. The app was designed to assist learners by providing information on their learning goal progress and to identify areas needing review by continuously processing measurement data and providing personalized instructional recommendations, all delivered in an anytime/anywhere mobile experience.

The goal of this app is to develop an XLAS-enabled prototype that supports learning through personalized recommendations and through free agency given to the students. The app also provides students with navigational opportunities to explore their career interests. In this section, we will describe how computational psychometrics principles are used to guide the development of this XLAS.

The app leverages multiple sources of data from ACT's portfolio of learning and assessment products. Beginning from its college readiness assessment, the ACT, the app identifies the underlying links from learners' measurement data to the taxonomic skills in English, Math, Reading and Science such as those defined by the HF (see Section 0.1.1), but is general enough to be applicable to any taxonomic classification of skills. The app gathers additional academic skills evidence from a workforce skills assessment (Applied Math, Reading for Information, Locating Information) where available. Beyond the core academic skills, the app evaluates Social Emotional Learning (SEL) data from the learner's SEL assessment results, that is the results from ACT's Tessera test. Blending these data, the app generates analytics that can predict mastery of skills at multiple levels in a taxonomy such as the HF.

Through the alignment of instructional content to taxonomic structures conducted with ML methods, the app is able to identify recommended resources to drive learning activities. The app makes targeted recommendations for learners at any selected or prescribed level in a taxonomy. It uses its knowledge of the learner's predicted abilities along with the understanding of hierarchical, parent/child relationships within the content structure to produce personalized lists of content.

With additional practice activity (e.g. like that found in test preparation quiz/test sessions) the app is able to continue to update and refine its predictive analytics and adapt its recommendations to learners over time. The app uses a clearly presented, three star rating system for the top level areas of a taxonomy (e.g. subject, domains) to communicate achievement to the learner and to encourage and highlight the next areas for review.

The Companion App also features access to navigational tools that were developed by ACT researchers, e.g. Cruce and Mattern (2018) These tools provide learners with insights about their career interests and the relationship of their personal data (e.g. assessment results, grade point average) to potential areas of study in college based on longitudinal, higher education outcome studies.

The Companion App was piloted with group of Grades 11 and 12 high school students in Clinton, South Carolina, USA between Fall 2017 and Spring 2018. The results were presented in an unpublished report (S Polyak et al. 2018)

0.5 Enabling a Personalized Learning Experience: Recommendation and Diagnostics (RAD) API

The development of the ACTNext Companion App led to the insight that the underlying capabilities could be packaged into an application programming interface (API) that can offer the diagnostic and recommendation engines as a service to other products or platforms. ACT Academy, built using a collection of free online resources from OpenEd, is one such system, a free online platform designed to help students master the skills they need to improve their test scores and succeed in college and career. Students can take authentic practice quizzes and receive personalized video lessons, games, and interactive resources whenever they miss a question. ACT Academy provides a free personalized path for students to help them prepare for the ACT test. Instead of teaching to the test and going through similar questions, ACT Academy videos help students learn the actual concepts being addressed in the questions. ACTNext defined a set of API methods that:

- accepts measurement data for a learner using a well defined, open standard, IMS Global's Caliper for learning events. The Caliper AssessmentItemEvent format is used to identify who the learner is, which item they responded to and the dichotomous outcome of their interaction.
- ingests the definition of hierarchical standard definitions using the IMS Global CASE standard. The CASE standard provides a machine-readable expression of the individual standard statements and conveys the hierarchical relationships (e.g. subject, domain, strand, sub-strands, etc.)
- automatically caches instructional content resources from any learning object repository (LOR) using the IMS Global LTI resource search API. By caching the LOR content that has been aligned to the standards, RAD can efficiently build personalized lists dynamically with up-to-date diagnostic data.
- continuously tracks learners' mastery of taxonomy skill/skill areas using modular, configured algorithms. As learner evidence is processed, RAD uses the algorithm that has been configured for the LAS to update new predictions of skill mastery.
- offers diagnostic engine results for learners based on any collection level in the taxonomy. RAD method calls allow for the parameterization of requests that can indicate an area of interest with respect to the taxonomy, e.g. return current estimates of Geometry skills within Math.
- offers recommendation engine generated, personalized lists of instructional content for any collection level in the taxonomy. RAD method calls allow for the parameterization of requests that can indicate an area of interest with respect to the taxonomy, e.g. return personalized resources for Geometry within Math.
- allows human curators to add content rules that require selected content for specified taxonomy areas. If human curators want to promote known resources for an area they can boost the importance of selected instructional content items at any level of the taxonomy (e.g. subject, domain, strand, etc.)
- permits bootstrapping of diagnostic estimates using prior assessment results. This feature allows learners to present results from a prior assessment (e.g. ACT score category report ratio data) in order to estimate skill knowledge prior to taking LAS assessments.

All of these methods are delivered from the secure, highly scalable, cost-effective, manageable platform, Amazon Web Services (AWS).

The initial integration of this API was completed in September 2017 with ACT's free test preparation web-based platform, ACT Academy. ACT Academy uses a customized version of the TAO open source test delivery platform capable of generating the IMS Global Caliper events (e.g. AssessmentEvent, AssessmentItemEvent) that the RAD API uses to continuously track

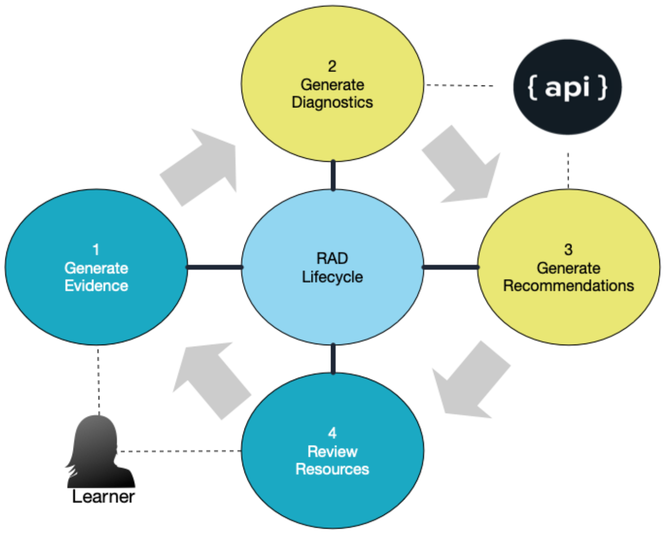


Figure 4. ACTNext Recommendation and Diagnostics (RAD) API

learner data. ACT Academy's dashboard/progress areas use RAD API's diagnostic methods to present learners with the same star-based progress report that was described above in the Companion App.

As learners navigate to the resources tab in ACT Academy, they are presented with personalized lists of instructional content that use their RAD diagnostic record to select the most relevant content based on the learner's needs. RAD effectively links assessment content and instructional content via the configured taxonomy while generating useful insights and analytics to help learners navigate their test preparation and skill practice goals.

We refer to this continuous cycle of activity between the learner in an LAS and the RAD API as the RAD life-cycle as shown in Figure 4. Learners generate evidence via assessment data which RAD then uses for its analysis. The analysis itself uses techniques from computational psychometrics. Essentially RAD connects low level evidence data up to high level estimates of learner abilities by applying algorithms (see Section 0.5.1), coupled with an item response model, to take into account a learner's prior estimate of ability alongside the system-derived estimate of skill difficulty. Other examples of the application of computational psychometrics for this project involve the machine learning techniques used to ascertain taxonomic tagging of resources and items. Today, a semi-automatic process is used to suggest tags to human curators that can then confirm or reject proposals, helping the system to better learn the classifications in the future.

The initial results of applying this computational psychometrics solution to learner diagnostic tracking has demonstrated that this is an important tool supporting our vision for unifying learning, assessment and navigation. We are currently expanding our approach to incorporate performance metrics that will:

- report continuous classification accuracy, i.e. how well is RAD predicting that learners would get items correct/incorrect based on its diagnostic data?

- use additional Caliper events such as MediaEvents to measure platform usage learning analytics, e.g. how many RAD recommended resources are learners reviewing?
- evaluate the fairness of the algorithms, by investigating the population distributions of star ratings and recommendations for an LAS. We want to provide aggregate analytics that show e.g. how many 1,2,3 star ratings have been made for specified populations.

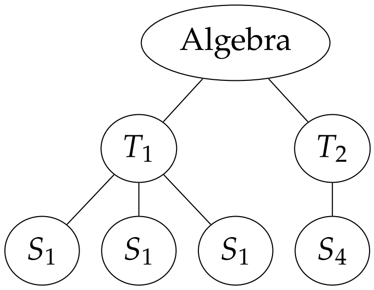
0.5.1 Diagnostic and Recommendation Models

To further detail the theoretical models, we focus next on psychometric models. Traditionally models for assessment relied on unidimensional models of latent ability. Such models are built to be able to correctly rank a set of learners from highest relative ability to lowest relative ability. However, these unidimensional models, such as models in Item Response Theory (IRT), are unsuitable for determining the source of these differences in ability. In other words they are incapable of diagnosing the underlying skills which the learners have or are lacking. Cognitive Diagnostic Models (CDMs) are built specifically for this purpose. Rather than modeling a unidimensional latent ability, CDMs equip each learner with a multidimensional latent variable where each dimension corresponds to a particular skill.

One particular project that we are working on and which utilizes concepts from CDMs is the Recommendation and Diagnostics (RAD) engine delivered through an API that was described above. The RAD API defines how a learning and assessment system can interact with the RAD engine. The RAD engine is built to be able to continuously track and update the skills in some hierarchical skill taxonomy, such as the ACT Holistic Framework. In ACT Academy a learner is able to choose a category from the Holistic Framework (HF; see Section 0.1.1) to practice. The skills in the HF are organized in a hierarchical tree structure called a knowledge graph. The knowledge graph is composed of a set of nodes along with their direct relationships represented by edges between the nodes. For each node k in the knowledge graph a learner has a proficiency value or profile value π_k, representing the probability that the learner has mastered this node. Together the knowledge graph along with a specific learner's proficiency values is called the Personal Learner Knowledge Graph (PLKG). The RAD API models the bottom most nodes (leaf nodes or nodes without any children) of each learner's PLKG. Estimates of a learner's proficiency in these bottom nodes is then percolated up the tree by averaging. See Figure 5.

The learners in ACT Academy see a relatively high level part of the HF known as the reporting category. Although the probability of node mastery is stored internally as a value between 0 and 1, it is presented to the learner in a discretized fashion as a star value (between 1 and 3 stars). Based on the learner's interests or evaluation of their HF mastery, as presented by the RAD API, the learner selects one of these categories to practice. They are then given a short set of items in the form of a quiz. The RAD API processes these responses in real time, updates the mastery of the leaf nodes, percolates the information up the HF tree, and the whole process is repeated again when the learner selects a new topic to practice.

From the above high-level overview of the RAD API diagnostic framework we can see that the statistical model underpinning the diagnostics and the algorithm used to update the parameters in the model needs to have several key features



$$\pi_{T_1} = \frac{1}{3}\sum_{i=1}^{3}\pi_{S_i}, \qquad \pi_{T_2} = \pi_{S_4}$$

$$\pi_{\text{Algebra}} = \frac{1}{2}(\pi_{T_1} + \pi_{T_2})$$

Figure 5. An illustration of the Personal Learner Knowledge Graph (PLKG) along with the process of percolating information up the hierarchical tree. T_1 and T_2 represent two topics under the category of Algebra. $S_1, \ldots, S_4$ are the skills that are estimated by the RAD API, where π_k represents the proficiency the learner has for skill k. The right hand side of the figure explains how information from the estimated skill or proficiencies percolate up the tree.

1. The algorithm needs to be able to process data in real time, updating learner profiles after either every response or after a small set of responses have been accumulated.
2. The model needs to be able to quantify a learner's mastery of every leaf node in the HF (and convert this into a value between 0 and 1).

Additionally, we require that one learner's activity should not change another learner's profile. Ideally the model would only have parameters that are updated in real time, however we will allow some hyper-parameters that are considered fixed but allowed to be updated on a longer term schedule with a batch process of the data.

Several algorithms and models were studied for the diagnostic engine. One of the algorithms selected to power the diagnostic engine of the RAD API was the Elo algorithm. First we describe the basic Elo algorithm and then we present how we adapted the Elo algorithm to fit into the RAD API. The Elo algorithm was developed to track and calibrate the rankings of players in competitive games. Its origins are from competitive chess (Elo 1978) An example of the use of the Elo algorithm in an educational context can be found in Pelánek (2016) The model underlying the Elo system is quite simple. Every player is associated with rating θ_p. The probability that player p beats player q is modeled as follows.

$$E_{pq} = \text{Pr}(p \text{ beats } q) = \frac{1}{1 + e^{-(\theta_p - \theta_q)}} = \sigma(\theta_p - \theta_q)$$

After every match the parameters for the players involved are updated taking into account the probability that player p wins and the actual outcome of the game.

$$\theta_p^* = \theta_p + K(X_{pq} - E_{pq})$$
$$\theta_q^* = \theta_q - K(X_{pq} - E_{pq})$$

where θ^*_p and θ^*_q are the updated values, X_{pq} is the result of the game (1 if p betas q and 0 otherwise), and K is a scaling factor.

For the RAD API we adopted a similar algorithm to estimate the values in a model that is inspired by the multidimensional Rasch model and the log-linear test model (LLTM) (see Pelánek (2017) and Pelánek (2016)). From the LLTM we use the idea of deconstructing the item difficulty to be linear in the difficulties of each skill. This allows us the ability to directly track skill difficulty. Let X_{si} be the response of student s to item i. We model the probability of a correct response as specified below.

$$E_{si} = \text{Pr}(X_{si} = 1) = \sigma(m_{si}) \tag{1}$$

$$m_{si} = \theta_s + \sum_k q_{ik}\theta_{sk} - \sum_k q_{ik}\beta_k \tag{2}$$

where θ_s is a general measure of ability of student s, θ_{sk} is a skill-specific measure of ability for student s on skill k, q_{ik} is a vector of 1's and 0's corresponding to which skills item i is tagged to, and β_k is the difficulty of skill k.

The update formulas are adapted from the original Elo algorithm to correspond to this model as as follows.

$$\theta^*_s = \theta_s + \frac{a_s}{1 + b_s \times n_i}(X_{si} - E_{si})$$
$$\theta^*_{sk} = \theta_{sk} + \frac{a_{sk}}{1 + b_{sk} \times n_{sk}}(X_{si} - E_{si})$$
$$\beta^*_k = \beta_k - \frac{a_i}{1 + b_i \times n_k}(X_{si} - E_{si})$$

where θ^*_s, θ^*_{sk}, and β^*_k are the updated values, $n_{\bullet}$ are hyper-parameters controlling the sensitivity (speed) of the values' update, n_s is the number of prior items answered by student s, n_{sk} is the number of items on skill k that student s answered, and n_k is the number of students that have answered an item utilizing skill k. Initial values for all parameters θ_s, θ_{sk}, and β_k were set to 0. Once the parameters of the model have been updated the profile values, π_{sk}, of the learner's PLKG are updated as follows.

$$\pi_{sk} = \sigma(\theta_s + \theta_{sk})$$

After which the rest of the learner's PLKG is updated by percolating the information up the tree. This blend of a psychometric model with an algorithmic/rating system is being described in Yudelson et al. (2019)

Discussion and Conclusions In this paper we outlined a comprehensive holistic learning and assessment system and indicated how the computational psychometrics paradigm integrates all these complex pieces. We discuss how learning, assessment and navigation need to be developed together to enhance the students' opportunities for a successful, holistic educational experience. A holistic learning and assessment system has many interdisciplinary components in which each individual component is an area for research and development: from the design, to data structures for big data, to mobile platforms, recommendation engines, the development of APIs and psychometric and ranking models for learning. Each of the areas described here include innovations, or at least extensions, of existing capabilities. Several papers are now being written simultaneously where the details of these approaches and their evaluations are being presented.

While significant progress has been made on the research and development of holistic learning and assessment systems, more work is needed to refine the methodologies, to continuously evaluate them for fairness, efficacy, and validity, and to scale them up. For example, the RAD API has been live for the past three months and guided students' choices for about 100,000 students so far; research on its efficacy and validity is in progress. The goal is to be able to provide all learners with access to quality educational resources and feedback, regardless of their background and geographic location.

Future research of XLAS includes the development of new types of dynamic cognitive diagnostic models that are appropriate for learning and of artificial intelligence (AI) and multimodal analytics to enhance these psychometric models. We also need to continue to work on more accurately aligned testing instruments and instructional resources via taxonomies.

For the authors, one of our future projects includes refining and enhancing the scalability of the Companion App. We are also developing additional micro-services to support multiple ways of personalizing and adapting the learning environment. Our goal is to integrate our research and prototypes with LAS partners and researchers, and extend the current XLAS work beyond the authors' organization with our Software as a Service (SaaS) model.

* Conflict of Interest Statement

All authors (Alina A. von Davier, Benjamin Deonovic, Michael Yudelson, Stephen T. Polyak, and Ada Woo) are employed by ACT, Inc.

* Author Contributions All the authors contributed equally to this paper. This is an invited paper based on the Keynote Address given by AvD at the International Test Commission, in Montreal, Canada in July, 2018. AvD, SP, and AW contributed to the conception and design of the work. BD, MY, and SP contributed to the acquisition, analysis, and interpretation of data. All authors contributed to the writing.

* Acknowledgments The work described here covers several large projects with many researchers and developers; in particular, we acknowledge the contributions of Gunter Maris to the models and recommenders presented here and of Kurt Peterschmidt to the development of the capabilities and mobile application. The authors thank Maria Bolsinova and Lu Ou for their feedback on the

previous versions of the paper, to Andrew Cantine for editorial help and to Matthew Livaudais for help with the graphics.

* References

Aguilar, S J (2018) Learning analytics: At the nexus of big data, digital innovation, and social justice in education. *TechTrends*, *62*(1) 37–45.

Arieli-Attali, M Ward, S Thomas, J Deonovic, B E & von Davier, A A (2019) The expanded evidence-centered-design (e-ecd) for learning and assessment systems: A framework to incorporating learning goals and processes within assessment design. *Frontiers in Education*. (*Under Review*).

Bobek, B L & Moore, R (2017, May) What can colleges do about the concerns of diverse college-bound students? In *Association for institutional research forum*. Washington D.C

Camara, W O'Connor, R Mattern, K & Hanson, M A (2015) *Beyond academics: A holistic framework for enhancing education and workplace success* (ACT Research Report Series No. 4) ACT

Cipresso, P Villani, D Repetto, C Bosone, L Balgera, A Mauri, M ... Riva, G (2015) Computational psychometrics in communication and implications in decision making. *Computational and Mathematical Methods in Medicine*.

Cruce, T & Mattern, K (2018) *Sticking to the plan: Which factors are related to intended-declared major consistency* (Working Paper Series No. 07) ACT

Deonovic, B Yudelson, M Bolsinova, M Attali, M & Maris, G (2018) Learning meets assessment. *Behaviormetrika*.

Doignon, J.-P & Falmagne, J.-C (1999) *Knowledge spaces*. Springer-Verlag Berlin Heidelberg.

Elo, A E (1978) *The rating of chess players, past and present*. Arco Pub.

Heffernan, N T & Heffernan, C L (2014) The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, *24*(4) 470–497.

Holland, J L (1958) A personality inventory employing occupational titles. *Journal of Applied Psychology*, *42*(5) 336–342.

Khan, S M (2017) Multimodal behavioral analytics in intelligent learning and assessment systems. In A A von Davier, M Zhu, & P C Kyllonen (Eds.) *Innovative assessment of collaboration* (pp. 173–184) Cham, Switzerland: Springer International Publishing.

Kruis, J (2018, December) A general framework for choice dynamics. In *28th interuniversity graduate school of pscyhometrics and sociometrics winter conference*. Arnhem, the Netherlands.

Mislevy, R J Steinberg, L S & Almond, R G (1999) *On the roles of task model variables in assessment design* (CSE Technical Report No. 500) University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

Paek, P L & Bobek, B L (2018, July) Unpacking the factors contributing to summer melt and impacting college readiness. In M Hailu (Chair) *College readiness and implications for student success*. Symposium conducted at the annual conference of the american educational research association. New York City, NY

Pelánek, R (2016) Applications of the elo rating system in adaptive educational systems. *Computers & Education*, *98*, 169–179.

Pelánek, R (2017) Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 1–38.

Polyak, S T von Davier, A A & Peterschmidt, K (2017) Computational psychometrics for the measurement of collaborative problem solving skills. *Frontiers in Psychology*, *29*.

Polyak, S Yudelson, M Peterschmidt, K von Davier, A A & Woo, A (2018, March 19) *Actnext educational companion pilot study report*.

Rosenberg, S (1992, April 12) Virtual reality check digital daydreams, cyberspace nightmares. *San Francisco Examiner*, C1.

Stoeffler, K Rosen, Y & von Davier, A A (2017, March 13) Exploring the measurement of collaborative problem solving using a human-agent educational game. In A Wise, P H Winne, G Lynch, X Ochoa, I Molenaar, S Dawson, & M Hatala (Eds.) *Proceedings of the seventh international learning analytics & knowledge conference*. Vancouver, BC, Canada.

Suskie, L (2018) *Assessing student learning: A common sense guide*. Jossey-Bass.

Teasley, S D (2017) Student facing dashboards: One size fits all? *Technology, Knowledge and Learning*, *22*(3) 377–384.

Tomlinson, M (2004) *14-19 curriculum and qualifications reform: Final report of the working group on 14-19 reform*. Department for Education and Skills (DFES)

van der Linden, W (2018) *Handbook of item response theory*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. CRC Press.

von Davier, A A (2015, July 10) Virtual and collaborative assessments: Examples, implications, and challenges for educational measurement. In F Bach & D Blei (Eds.) *International conference on machine learning*: *Workshop on machine learning for education*. Lille, France.

von Davier, A A (2017) Computational psychometrics in support of collaborative educational assessments. *Journal of Educational Measurement*, *54*(1) 3–11.

von Davier, A A Wong, P Polyak, S & Yudelson, M (2019) The argument for a "data cube" for large-scale psychometric data. *Frontiers in Education*. (under review).

von Davier, A & Arielli-Attali, M (2018) Extending the ECD to navigation. Unpublished notes Work in progress.

Whitmer, J Nasiatka, D & Harfield, T (2017) *Student interest & patterns in learning analytics notifications*. Blackboard Analytics.

Wiswall, M & Zafar, B (2015) Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, *82*(2) 791–824.

Yudelson, M Deonovic, B Chopade, P & Polyak, S T (2019) Towards dynamic adaptation and personalization in act academy – a free online learning platform. In M Bolsinova (Chair) *Measurement in adaptive learning systems: Challenges and solutions*. to appear. Symposium conducted at the annual conference of the national council on measurement in education.