

4

Assessment Futures

There's a fascinating filament woven into the fabric of *Star Trek*: throughout many of the various television (TV) shows, films, and spinoff transmedia content, the utopian vision of science-driven exploration and peacekeeping builds on all-too familiar assumptions of academic evaluation.

Briefly, *Star Trek* is perhaps one of the most historically consistent, though flawed, set of perspectives on just futures in US popular television culture. As Jean-Luc Picard, captain of the *USS Starship Enterprise*, suggests, "The acquisition of wealth is no longer the driving force in our lives. We work to better ourselves and the rest of humanity." Popular media has rarely so succinctly or radically stated such a fundamental statement of values: the embrace of betterment, which includes antiracism, learning, and education. Apocryphally, *Star Trek* was the only TV show that Martin Luther King Jr.'s family would watch together, and he famously saved the series by personally urging Uhura actor Nichelle Nichols to stay on the show (NPR, 2011). Despite such a powerful endorsement, the series remains a reflection of its time, and its treatment of "betterment" is often as sexist, racist, and homophobic as the much less thoughtful TV shows that ran alongside it. Considering the flawed but uplifting space that this show

has maintained for more than half a century, the way it treats schooling in myriad episodes is both well-articulated and entirely retrograde.

Watching the 2009 film reboot, we were surprised to encounter the strict representation of Spock's educational regimen on Vulcan, often portrayed as being "past racism." With their ideological commitment to logic above all else, this interpretation is, unfortunately, not only a reification of dominant educational strategies of US schools in 1997, but also a series of assessment-based schooling principles built on a "logic" of antiblackness and a culture of poverty. Clearly, the movie made no connection between punitive "drill and skill" flashcard methods of early educational technology and "illogic," though we have long known that these methods do, effectively, nothing to teach or measure understanding. Indeed, education, as envisioned in the film, still involves racist bullying (against Spock, for being part human).

One thing that is curious about this development within the film, and across portrayals of education in *Star Trek*, is that the offered academic "AI tutor" is curiously unintelligent. Lacking the nuance of irony, humor, or even the basic flexibility to extend or imagine what a character is seeking, these tutors are portrayed as rote, inflexible, and unengaging. In a world with widespread holodecks, and the possibility to create artificial intelligence smarter than humans, the tutor is simply glorified flashcards. Perhaps even more improbably, this treatment of the old as an assessment-driven new is entirely in line with academic approaches to learning in the non-Starfleet-guided world of the present day.

Perhaps the most infamous assessment at the heart of *Star Trek*, the Kobayashi Maru test, is a standardized exam that is meant to reveal an individual's character based on an unwinnable scenario; it is an adaptive assessment rooted in failure. The Kobayashi Maru test is a simulation in which failure is intended to be inevitable, designed to assess a potential Starfleet captain's resilience in the face of catastrophic failure. While Captain Kirk is diegetically famous for being

the only Starfleet student ever to pass the exam, the 2009 film depicts his “win” as done essentially through hacking the system’s coding and bypassing the pedagogical lessons related to encountering failure. The Kobayashi Maru assessment, like the rest of the depictions of educational evaluation in the distant future of *Star Trek* boldly stays rooted in the stale and unimaginative present.

Considering a failure of imagination when it comes to assessment and schooling in the hopeful articulations of white-authored science fiction, this chapter begins with a foundation of how assessment falls across the AnSpec before offering alternative examples to drive research design and practitioner implementation.

What Is Assessment?

Assessment is a term that pretty much everyone in educational research dreads. To a principal, it might mean verifying that classes are teaching some specific content. To a teacher, it might mean a test. To a student, it might mean judgment. To most, the word itself is more of a threat than a specific class of ideas. In the abstract, it simply means evaluative information; people receive evaluative information as *feedback*. What happens with that feedback and how an individual can respond and act upon it, are important outputs of feedback and assessment models, necessary for research design to constantly keep in mind. At the same time, feedback is a mixed term. It may squeal dissonantly on a guitar. It may mean that people think your pants look nice (or not). It may be helpful guidance on what you can do better next time. In almost every model of learning—cognitivism, behaviorism, direct instruction, sociocultural constructivism, apprenticeship, and so on—feedback plays a central role. Indeed, it is almost impossible to imagine a version of improving or augmenting knowledge or processes without some evaluation of current knowledge or processes.

Feedback and assessment are at the center of most traditional models of learning and of processes for growth and development writ large. We learn through feedback measures, and these become systematized and made into routines of schooling, for better or, more often, for worse. Such feedback typically helps evaluate measures of competence and ranking measures. The SAT and ACT are, fundamentally, ranking measures. One's score only has meaning relative to those of other students, and the score is either higher, lower, or the same.

Though the assessments that may come to mind (and are exemplars in parts of this chapter) are the big, high-stakes, anxiety-inducing summative tests that may function as gatekeepers in different parts of a person's individual growth, assessments often function in formative ways. Measuring progress continually and along the way to the culmination of a learning journey is an important component of education design and this chapter's consideration of assessment design, measurement, and adaptivity holds for both summative and formative models.

Further, assessment models are not static. They are often developed to adapt to the needs of test designers, takers, and readers of the results. Although we will consider adaptive assessment as a core premise throughout this chapter, we offer two sets of design questions for teachers, researchers, and policymakers based on adaptive assessment:

- *How* does assessment adapt in your design and implementation? What are the flexible components that undergird adjustment?; and
- *Who* does this adaptivity address? What stakeholder needs are taken up in your adaptivity design and why?

These are not questions that have been pushed imaginatively within schools and educational research contexts. Historically, adaptation is tied to market-driven demands and to outputs related to college and career readiness. Adaptation is often understood as related

deficit views of learners—linguistic, cultural, and age-based framing of “who individuals are”—rather than framing their potential capacities in other ways. Likewise, some form of adaptive assessment is almost always included in the promise of an equitable and transformative future of education. Adaptive assessment is, at its core, the idea that a learner will be evaluated by their practice—that an evaluation of someone’s understanding of a topic will change to fit the person. Yet, when it comes to how assessments are understood in educational policy, in on-the-ground enactment, and in the public consciousness, tests usually are greeted with a mixture of drollery and dread. Assessments can feel boring, punitive, and unfair.

There is little to no qualitative feedback embedded in a numeric test grade out of 100. That fact is central to the logic of quantitative assessment, making it easier to abstract and compare student performance. The driving test, in comparison, is a test of competence: you either pass or fail. Passing is the result of the testers deciding that you are sufficiently skilled at driving; failure means the opposite. There is no concept of ranking; it does not make sense. Similarly, few assessment scholars would argue that someone who has done well on the SAT is skilled at any task or has any particular understanding other than doing well on the SAT.

Assessment and Data

As we explore the role of assessment, we need to ask a fundamental question: How has assessment gone so wrong? Consider most formal measures of assessment and testing that take place in schools and societal governance. From high school exit exams to SATs to Graduate Record Examinations (GREs) to tests for drivers’ licenses to bar exams to practice law in various states, formal assessments mostly fall into very specific genres. Such a limited approach to assessment clearly produces high stakes. Student outcomes and graduation rates

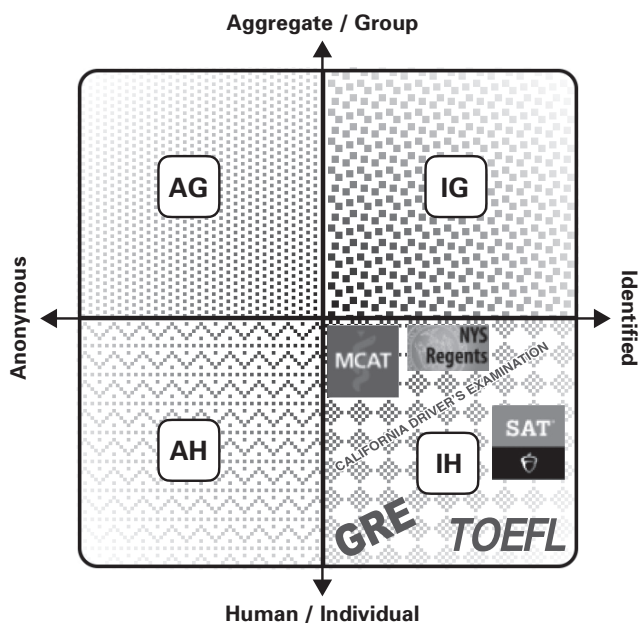


Figure 4.1

Formal assessments plotted on AnSpec.

and teacher evaluation measures in K-12 contexts are both prominent ways that assessments have lasting outcomes beyond measuring a skill set. Further, the cultural legacy of tests shapes outcomes profoundly. That pit in your stomach before a late-night cram session may, depending on your circumstances, reflect a legitimate rationale: stereotype threat, testing bias, and the gatekeeping of the profit-driven “test-prep” industry all speak to the unfair footing that economically-disadvantaged people have when it comes to taking tests.

We want to illustrate how assessments in school contexts typically fall across the AnSpec. This includes typical assessments encountered in schools, Scholastic Aptitude Tests (SATs), exit exams, Advanced Placement (AP) tests, Graduate Record Examinations (GREs), and so on. Plotting these different exams across the AnSpec looks like figure 4.1.

Looking at this figure, the pattern is perhaps all too obvious. All these examples, as well as most assessments that happen throughout K-12 schooling, fall into the Identifiable Human (IH) quadrant of the AnSpec. Though the data from exams might be parsed and examined in different research contexts, from the vantage point of the test taker and of the assessing entity (a teacher, organization, etc.) these tests all function in near-identical ways on the AnSpec: a test is taken by an individual whose identity is recognized by formal measures. At the risk of stating the obvious, there are good reasons for this pattern in schooling assessments. It is easy to locate where a specific individual's progress or proficiency falls across expected measures if we know who that person is as an individual, especially if we can compare that person to groups of similar people. However, there are also good reasons *not* to reinforce this pattern of measuring and gatekeeping within society.

What can we do? Why can we not imagine other possibilities for how assessment is enacted or developed in schools? It's not that educators love the IH-centered approaches to assessment. We posit that this clustering in purpose and design is something largely invisible to how our field designs and enacts assessment.

We recognize that assessment and evaluation are giants of the educational research realm. They are (often) well-funded, rigorous disciplines that guide student and teacher learning experiences. This chapter does not seek to uproot (all of) the practices and research-based knowledge about educational assessment or teacher evaluation. Rather, we hope that the spread of approaches cast across the AnSpec will allow our field to point to and wonder about unexplored possibilities in assessment design. *What would happen if we repositioned student identity? What if our evaluation of student knowledge was adaptive across different dimensions?*

As we move across the AnSpec quadrants to explore adaptive assessment, throughout the rest of this chapter, we offer some guiding

Table 4.1

Adaptive assessment questions for purposeful design

- Why do formal schooling approaches to assessment (adaptive or otherwise) cluster in specific quadrants? What does this say about imagination and educational design?
 - How do the clusters of assessment in the past and in your own designs speak to the treatment and categorization of student knowledge and interest?
 - What kinds of agency does an assessment presume within the individuals taking it?
 - How could various kinds of learning spaces move fluidly across AnSpec? What do such spaces require for powerful design?
-

questions about both the past enactments of adaptive assessment and about the role of design, research, and speculation for future-oriented designs (table 4.1).

Playful and unexpected examples of adaptive assessment nestled in everyday life may at first appear like benign diversions from the august world of school-based assessment. We offer playful approaches to assessment to intentionally seek data that is rooted in justice. By (hopefully) upending the perceptions of what counts as an assessment and how such data is developed, collected, and used, we are attempting to extricate assessment from its singular and limited residence in the IH quadrant.

Speculative Assessment, Data, and Justice

Given this chapter's goal of exploring a wider perspective on adaptive assessment, examples of adaptive assessment can come from software in which it serves a transparent purpose. In playful contexts, many games use some form of adaptive assessment, explicitly or implicitly. Even board games often include complex mechanics to “even the

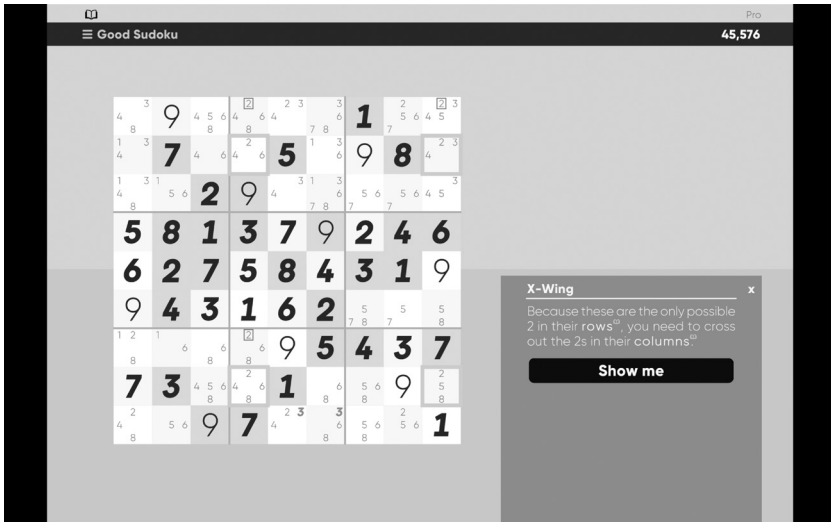


Figure 4.2

Good Sudoku (Gage, 2020) suggesting a possible course of action.

playing field” in ways that require an “assessment” step to evaluate current success.

Anonymous Grouped (AG)

One particularly thoughtful example of a public-facing puzzle game is *Good Sudoku* (Gage, 2020). Sudoku (figure 4.2) is a number game played on a 9×9 grid in which you want to find the single unique solution in which the digits 1–9 are not repeated on a row, a column, or within a “house” (a 3×3 subdivision of the board).

Good Sudoku is frankly a terrific game for several reasons. One of its core features is that it assesses the quality not only of your play (i.e., score) but also of your strategies, using a relatively complex algorithm (described in Eggplant: The Secret Lives of Games, n.d.) that is both clear that it is being used but is designed to feel “unobtrusive” to the player. Over time, the adaptive assessment finds examples of times that you must use a specific named strategy to make a play on

the sudoku board. If you correctly use the strategy a predetermined number of times, you are assessed as being proficient, and a pop-up announcement informs you that your competency has been evaluated. If you make a mistake with that strategy multiple times, it suggests specific training in that strategy. Also, you can ask for hints, and it will explain to you why a given strategy is the correct play for your specific board.

This is AH data: it does not care who you are, other than “someone who would pay to play sudoku,” but it records every relevant play and makes judgments on how best to support your progress. The only negative feedback you might receive is nothing more than a lower score in a game where scores are all but meaningless. That is, if the player eventually solves the board correctly, scores in sudoku are more contingent on the semi-random configuration of the board than the skills of the player. Most of the other sudoku games we’ve played do not even keep a score other than “boards solved per difficulty level.” It may be that the “noninformationality” of the score is what leads to the possibility of more robust feedback in assessments.

The inversion of that statement becomes: “Informational scoring likely leads to less robust assessment feedback.” This refers to situations in which the scoring is plain and the score itself provides the correct answer. For example, if a teacher marks the statement “All elephants have polka dots” as incorrect on a test, the graded student now knows that some or all elephants do not have polka dots. Very little else is needed or given, but there is also no intuition that this information will be used in any contextual way.

Identifiable Grouped (IG)

Matthew has spent several years working on the *Beats Empire* project with scholars, researchers, designers, students, teachers, and schools to create a “just” assessment that lives primarily in the IG space. *Beats Empire* is designed to be a *culturally relevant assessment game* for New York City (NYC) public schools that provides feedback to

teachers regarding what their *class* understands about the grade-level computer science content. In *Beats Empire*, students manage a music studio and collect and use data to make decisions about signing new artists and recording new songs. There is no single goal: students can choose from a variety of aims, including maximizing money, reputation, and/or fans, or simply making auto-generated music.

The idea is that students play the open-ended game freely just to have fun. Although there are assessment opportunities built into the game, they are entirely unforced. An individual student might play the game across multiple days of class and remain “unassessed” on a given topic. This is very much on purpose: school, by default, assumes individualized (or small-group-level) assessment; there is very little classroom-grouped assessment on which feedback is given to the students and teacher(s). In *Beats Empire*, it is literally impossible to assess a student negatively. There is no way to enforce the “hitting of assessment targets,” so it is much more likely for an individual student to avoid encountering a given assessment opportunity than for that student to fail to understand target content. That said, in each class, a subset of students will almost certainly encounter all assessment opportunities, which provides the teacher information about what the students, in aggregate, do or do not understand.

An open question remains as to how one can judge or support “cultural relevance” in situations in which the students are not identified. In the case of *Beats Empire*, Prof. Betsy DiSalvo (Georgia Tech) led a team of researchers, students, and teachers in co-design sessions in NYC public schools. In addition, we did various levels of fact-finding about both student-content connections and content-theme connections. Though the initial model was not necessarily music-industry-themed, the resulting game leaned heavily both on students’ descriptions of the music industry and the computer science behind analysis in the music industry. The music was one of very few cultural touchstones that every student with whom we talked (out of hundreds in total) had both interest and familiarity; as an industry,

the “moguls” are remarkably diverse and do not fit any easy categorization. In doing data analysis, however, we often found ourselves butting up against the other side of that coin: students would spend a lot of time customizing the game with the music genres they preferred. This worried us: perhaps enough customization would speak to individual student identity? In checking, we did not find any systematicity to who did this and the choices they made; however, this lack of systematicity may speak more to the ways that music is taken up in NYC than to anything else.

The potential upsides to the project seem clear: students enjoy learning and interacting with the target content; teachers get feedback about which content their classroom understands best; there is little opportunity to rank, punish, or judge individual students; there is less of a “privacy surface” for any individual’s data; and the teacher remains the authority on teaching and assessing their classroom. The biggest tradeoffs, then, become clearer: it is a deeply unfamiliar modality of assessment. Indeed, the idea that the assessment exists at the classroom level has been something that many teachers have had difficulty integrating. Historically, if a student has had trouble understanding an idea, a teacher can work with them on it. At the classroom level, do “problems interpreting bar graphs” imply that the syllabus should be amended? What if you, as a teacher, believe that the students know otherwise? What if they have recently done pretty well, in aggregate, on an exam that used bar graphs?

These issues led us, for better or worse, to a teacher-focused dashboard tool rather than a report. Our dashboard (Kumar et al., 2022) supports live classroom management at the level of “Bobby has been idle for 18 minutes”; offers classroom “reports” on what is clearly understood; and provides “insights” with some suggested options for what to do about the reports. For instance, if nobody in the class has used “digital storage” in the game and time is limited, the teacher could suggest the value of using it; conversely, if most students have

used digital storage unsuccessfully, it may be time to have a quick refresher lesson on the relationship between data collection, storage, and access.

Fundamentally, these matters bring up the classic problems of any assessment: What does an assessment tell us? What is being assessed? How well does this assessment map to future tasks or future assessments? These questions are, perhaps, even more difficult to address when the data are not individualized.

Identifiable Human (IH)

In the opening of this chapter, we explored the clustering of multiple school-based assessments as Identifiable Human (IH) data in traditional education. As we saw, many people believe it is preferable or even definitional to education in the US. It is possible to imagine many situations in which you would like the data collected to be individualized and identifiable (that is, IH). *Personalized learning* is a term often used to envision a better future for education, in which each student is taken seriously as an individual. Strengths and weaknesses are evaluated individually and are not bound by the speed or skill of other students, such that a student might be able to progress without public embarrassment or fear of failure. Such a system might provide the opportunity to try out ideas with a personalized tutor in a more pedagogically sound manner. The science fiction series *Ender's Game* posits this sort of personalized world in which each student is individually tutored (and assessed) by an artificial intelligence (AI) and, as a result, individuals would not be held back by the nature of classroom teaching. There is a visceral strength to this vision: many people who try to envision educational futures often come up with models that involve personalized assessment or personalized instruction at some level (often exemplified by a robot such as, say, H.E.R.B.I.E. or H.U.B.E.R.T. in Marvel Comics; see Byrne, 1982). It happens so often in such similar ways in movies, TV shows, comic books, and novels that even a short list of these would drag.

The fundamental problem with these mechanized models is that many of these examples and models incorrectly suggest (or assume) learning to be an individual endeavor, one of the things we know to be false about human learning. Humans, to put it simply, learn better with and around other humans, even in situations in which automated tutors exist.

We did not always assume this about learning and assessment. Indeed, as we saw earlier in the chapter, most current schooling assumes that assessment should happen in isolation, though that model is intrinsically at odds with everything we know to be true of assessment.

Matthew received his Ph.D. from Northwestern University, one of the first “Learning Sciences” programs in the world. It was founded, in part, by Roger Schank, whose concept of the goal-based scenario (or GBS) was originally designed as a replacement for school. (Indeed, Schank repeatedly expressed his desire to “blow up the schools,” though that meant something different—significantly less violent—in the 1980s.) A fundamental conceit of the GBS was that people are all interested in *something*: some set of professions or tasks or experiences. The classic example from the GBS world is a firefighter. The GBS would teach that student all the content they needed to know to become a firefighter, and the assessments would be designed in concert with firefighters. However, because a GBS is computational, the content could be constructed, manipulated, and structured in ways that a learner would come to understand complex content in the context of firefighting. For instance, one would learn calculus because it is necessary to become a phenomenally skilled firefighter—perhaps pertaining to flow rates under heat and pressure. In the GBS, the traditional content of school that was not or could not be useful for a given profession would be ignored, and the assessments could theoretically be tailored to the individual student’s interests, skills, and talents.

In hindsight, Schank and his successors found the problems with GBS to be manifest and manifold: it assumes very difficult—if not

altogether false—things about how we might design for learning and assessment. Without pedagogues judging, assessing, and/or being part of the learning experience, people may learn little and experts trust the results perhaps even less. Furthermore, making an automated learning experience and automated assessment strategy for every possible set of student interests is an unfathomably huge task. There exist tens of thousands of possible contents with different sets of information—maybe you could break this idea down to a few hundred, but this is still a massive undertaking that would take tens of thousands of people many years just to verify, even if the simulations could somehow be generated. Schank, at the time, suggested that the modern institution of secondary schooling takes tens of thousands of people thousands of hours a year, but that we are confident that it teaches people relatively little.

Schank and Abelson's (1977) vision of assessment was that experts could come up with more authentic assessments of the kinds of tasks that they found valuable—but content expertise is not pedagogical content expertise, nor is it assessment expertise. Basically, being a terrific swimmer does not mean you are any good at teaching people how to swim, nor does it mean you are much better at figuring out how good people are at the subtasks that make up swimming (even if you might be very good at assessing if someone is an expert swimmer).

Anonymous Grouped (AG)

Anonymous grouped (AG) cases are simultaneously the most common and the hardest to describe. There exist innumerable situations in each day in which you are assessed both anonymously and in a group. In some sense, every randomized controlled trial, every election, every concert, every crowded event is an AG assessment.

In the US, voting is a canonical AG assessment: people are assigned representatives and referenda are passed based on the ways that people participate and act in groups. It is enshrined in law that these things are inherently anonymous. It is not possible in the US to find out how

an individual has voted. This is for good (probably obvious) reasons: if you were assessed as an identified individual, then the government you did not vote for could arrest you based on your vote. Sporting events feel like natural AG assessments. For instance, often players on a baseball team will do better when the crowd is louder. The crowd can give energy as a group to the players and the players will often assess the value of the crowd: “Their hearts weren’t in it tonight!” Similarly, concerts have a language for their assessments: “I can’t hear you, Toronto!” As with everything in this book, all these examples exist on a spectrum. A concert is purely anonymous and aggregated; it is impossible for a performer to tell if you, specifically, sang along, but easy to tell if “people are singing along.” Basically, the more the data can be aggregated, the more inherently anonymous they are, and the less chance there is for any possibility of identification.

The question remains: Are these examples, strictly, assessments? They are assessments in the sense that performance and action result in judgment, consequences, and feedback. They are not assessments in the sense that it is often hard, due to the level of aggregation, for a participant to evaluate their own participation from that feedback or judgment. That is not to say that AG assessments are meaningless, but rather that their meaning exists at a different grain size. Baseball teams make radical changes if people stop showing up to games or start cheering more quietly. Concerts will avoid “dead” towns. Bad politicians will be elected. These results all stem from AG assessments, and the implications—while at an aggregated level—affect all the individuals who participate in them.

Toward Just Futures: Radical Visions of Cross-Spectrum Assessment

This chapter began with the dreary reminder that assessments are rigidly mired in tired molds. In popular culture, in public imagination,

and in the lived experiences of young people today, all assessments seem born of the same mundane model that houses implicit bias and undue stress for test takers. When we look beyond schooling contexts, this simply doesn't have to be the case. When we imagine assessments that are more than the summative task of identifying a particular individual and declaring how that person measures up in the world, the possibilities of assessment—and the outcomes of it—become potentially transformative.

As designers of assessments, what might it mean to push our vision of design beyond singular and identifiable individuals? Like the previous chapters in this section, we think there is a necessary task in our field of playing across the AnSpec, dabbling in assessments that honor anonymity and imagining how such models might transform individual learning experiences.

Even in the future, the most radical visions of assessment that lurk in the flashback scenes of popular science fiction reveal digital flashcards and individual-assisted tutoring systems. We *already* have these models in the present day, and they thrive alongside generations of educational inequality. Missing in many of these models is the treatment of data: where do the results from assessments *go* when students take the SAT today or when Captain Kirk aces the Kobayashi Maru test in *Star Trek*? From our vantage point, there seems to be an impoverished view of how learners' data might be used ethically in the present and future visions of assessment and data science.

As we wrap up this chapter, let's go back to *Star Trek* for a moment. Remember all those restrictive assessments and the regimented forms of tutoring systems and tests designed for failure? Fortunately, that's only one dimension of the projection of learning and evaluation in the future. From our vantage point, the storytellers in *Star Trek* seem committed to data privacy in surprisingly deep ways. Though networked computers very much existed at the time, private data generated by individuals—and even the tracking location of those individuals—was encoded physically onto disconnected tablet devices called

PADDs (sometimes credited as an inspiration for the Apple iPad). In one scene in *The Next Generation* series, Counselor Troi is moving out of her office, and she finds that she cannot cram enough PADDs into her suitcase. In other words, you cannot even move data between PADDs, and, if you lose them, those data are lost. In every series of *Star Trek*, as far as we can tell, anyone living in the United Federation of Planets (the “good guys” for whom the *Enterprise* is an exploration ship) can put effectively unbreakable encryption onto their private/personal data if they so wish.

The vision of data privacy hinted at in the PADD design is an important consideration for assessment designs and helps reframe a consideration of whose data is anonymous for whom. The AnSpec is not a tool for simply rigid plotting of different assessment batteries. Rather, we want this heuristic to help flex and push on the boundaries of what is possible in classrooms transformed from the here-and-now. Fundamentally, any vision of “adaptive” education involves personal data: how else would it “know” you well enough to get better if you don’t know what understandings you lack? Just as *Star Trek* cannot depict a world in which those data are used both fairly and privately, we see the tensions underlying identifiable and anonymous, between range and scale, and between grouped and individualized data at odds in assessment development. As we seek assessments that live *across* the AnSpec, this chapter illustrates possible pathways across spaces unfamiliar to traditional modes of schooling.

This is a section of [doi:10.7551/mitpress/14381.001.0001](https://doi.org/10.7551/mitpress/14381.001.0001)

The Left Hand of Data

Designing Education Data for Justice

By: Matthew Berland, Antero Garcia

Citation:

The Left Hand of Data: Designing Education Data for Justice

By: Matthew Berland, Antero Garcia

DOI: 10.7551/mitpress/14381.001.0001

ISBN (electronic): 9780262377645

Publisher: The MIT Press

Published: 2024

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

© 2024 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC-BY-NC-ND license.

This license applies only to the work in full and not to any components included with permission. Subject to such license, all rights are reserved. No part of this book may be used to train artificial intelligence systems without permission in writing from the MIT Press.



The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Berland, Matthew, author. | Garcia, Antero, author.

Title: The left hand of data : designing education data for justice /
Matthew Berland and Antero Garcia.

Description: Cambridge, Massachusetts : The MIT Press, [2024] | Includes
bibliographical references and index.

Identifiers: LCCN 2023030088 (print) | LCCN 2023030089 (ebook) |
ISBN 9780262547529 (paperback) | ISBN 9780262377652 (epub) |
ISBN 9780262377645 (pdf)

Subjects: LCSH: Education—Data processing. | Education—Research. |
Educational evaluation.

Classification: LCC LB1028.43 .B45 2024 (print) | LCC LB1028.43 (ebook) |
DDC 370.285—dc23/eng/20230718

LC record available at <https://lccn.loc.gov/2023030088>

LC ebook record available at <https://lccn.loc.gov/2023030089>