

# The Sorting Machine

## *Special Education Classification as Infrastructure*

Jake Lawrence · Independent Researcher

*Author's note: This essay was prepared for journal submission. All identifying details in the opening vignette have been changed. The case is a composite drawn from published research and practitioner accounts, not a single identifiable child. The analytical framework is the author's own.*

---

### **Abstract**

The special education disproportionality literature has organized itself around a question that seems obvious: are students of color correctly identified under IDEA disability categories? Decades of careful research have produced two well-supported, opposed answers (overrepresentation as misidentification on one side, underrepresentation after covariate adjustment on the other) and the debate between them is genuinely unresolved. This essay argues that the impasse is diagnostic. Accuracy analysis assumes disability categories are measurement instruments picking out a pre-existing condition. But a measurement instrument and an infrastructure are different kinds of things: an instrument measures a condition that exists independently of it, while an infrastructure produces the conditions it appears to describe. IDEA disability categories are infrastructure in the precise technical sense defined by Susan Leigh Star and Geoffrey Bowker: embedded in funding formulas, legal obligations, teacher training tracks, and assessment batteries that together constitute what “disability” means in a given school. Treating them as infrastructure rather than instruments shifts the central question from “are the rates right?” to “what does the system do, and what was it built on?”

*Keywords: special education, disability classification, IDEA, disproportionality, infrastructure theory, equity, race*

## **The Room**

The school psychologist's report runs to eleven pages. It begins with a referral reason, “significant behavioral disruption, possible emotional disturbance,” and proceeds through standardized assessment scores, behavioral rating scales completed by two teachers, a developmental history gathered from a single parent interview, and a summary of disciplinary records spanning eighteen months. On page nine, there is a checklist. The student meets five of the six criteria for Emotional Disturbance under the Individuals with Disabilities Education Act. The sixth criterion, that the condition is not primarily the result of intellectual disability or health impairment, is noted as “not applicable.”

Two weeks later, an eligibility meeting convenes. Present: the school psychologist, the student's general education teacher, a special education teacher, an assistant principal, and the student's mother, who has taken a half day from work. The team deliberates for forty minutes. The student is found eligible. An Individualized Education Program is developed, calling for placement in a self-contained behavioral support classroom for four periods per day, with supplementary services including weekly counseling and a behavior intervention plan.

The student is nine years old. He is Black. He will carry this classification, and the institutional arrangements it triggers, through the remainder of his K-12 education.

Nothing in this sequence is unusual. No one in the room acted in bad faith. The school psychologist administered valid assessments. The teachers completed the rating scales thoughtfully. The mother asked good questions. The team followed the process the law requires. The infrastructure worked exactly as designed.

*That is the problem this essay examines. Not the bad-faith application of a system, but the system's faithful operation.*

The question is not whether the team made an error. The question is what kind of system was operating in that room, how it was built, and what it was built on top of.

## **An Impasse and What It Means**

Special education researchers have spent decades asking whether the rates at which students like this child are classified under IDEA disability categories are accurate. The question seems straightforward. IDEA categories exist to identify students who need services. If students of color are classified at higher rates than their white peers, either those students genuinely need services at higher rates, or the system is misidentifying them. The debate has organized itself around these two poles.

On one side, researchers documenting racial disparities argue that Black, Native American, and Latino students are overrepresented in high-stigma, low-resource categories, particularly Emotional Disturbance and Intellectual Disability, and that this overrepresentation reflects systematic misidentification: implicit bias in referral decisions, culturally loaded assessment instruments, and a history in which disability categories were used to maintain separate schooling for students of color after formal segregation became legally impermissible (Blanchett, 2006; Skiba et al., 2016). If this account is right, current policy is harming students by placing them in stigmatizing, restrictive environments they don't belong in.

A body of research using matched-comparison methods and extensive covariate adjustment argues, to the contrary, that after controlling for socioeconomic status and academic achievement, racially minoritized students are actually underrepresented in special education services (Morgan et al., 2015). The apparent disparity, on this account, is an artifact of higher rates of genuine need, driven by poverty, under-resourced schools, and accumulated disadvantage. Reducing identification rates would deny services to students who need them.

Both positions are carefully reasoned and empirically supported. The debate acquired its most pointed recent intervention in January 2026, when Rachel Fish, Kenneth Shores, and João Souto Maior published a two-pronged critique of the underrepresentation literature in *Exceptional Children*, arguing that its covariate adjustment practices cannot distinguish between the effects of disability and the effects of educational inequality itself (Fish et al., 2026). When researchers control for test scores to isolate “disability” from “disadvantage,” they may be controlling away the very conditions that produce, and constitute, the need for services.

*They are not arguing about the facts. They are arguing about what the facts mean, because they are measuring different things with the same numbers.*

The debate is genuinely unresolved. And thirty years of careful, opposed research reaching irreconcilable policy conclusions from the same data is not simply a problem of insufficient evidence. The disagreement runs deeper than evidence: into the question of what the evidence should be *of*.

This is what a methodological wall looks like. And a wall is useful, because it tells you which direction to walk.

The accuracy debate asks whether the sorting machine is sorting correctly. The framework applied here asks what kind of machine this is, how it was built, and what it was built on top of. Those are different questions. The second set generates answers the first cannot reach. Not because infrastructure theory is superior to accuracy analysis, but because the specific pathologies of special education classification are infrastructure pathologies. They are properties of how the system was constructed, not primarily properties of how its operators perform.

## **What Infrastructure Asks**

The sociologist Susan Leigh Star and the information scientist Geoffrey Bowker spent two decades studying what happens when you take classification seriously: not as a philosophical problem about whether the categories are right, but as an empirical question about what the categories *do* (Bowker & Star, 1999). Their 1999 book, *Sorting Things Out: Classification and Its Consequences*, examined how systems for sorting (diseases, racial identities, nursing tasks) generate institutional

consequences that outlast and outrun the intentions of their designers. Classification systems are not neutral instruments imposed on a pre-existing reality. They are built things, with histories and politics and switching costs, and they actively shape the realities they appear only to describe.

Star condensed this into a framework with eight characteristic properties, from embeddedness to visibility upon breakdown (Star, 1999). The most important implication: infrastructure persists not because it is optimal but because the cost of replacing it exceeds the benefits of replacement at almost every moment in its history.

Treating IDEA categories as infrastructure is not an ontological claim about what those categories fundamentally are. It is a decision to ask infrastructure questions (what is this system embedded in, what was it built on top of, what does it cost to change, when and to whom does it become visible) because those questions tend to reveal things that accuracy questions cannot.

I've spent the professional portion of my career implementing and migrating classification systems in government and enterprise contexts, watching technically superior replacement systems fail to displace the ones already running, because the old systems are installed in every workflow, every training manual, every person's muscle memory. The infrastructure frame is not imported from an adjacent discipline. It is the frame that practitioners working inside classification systems already use, without always having the vocabulary for it.

What follows applies three of Star's eight properties, the three most productive for this domain, not to demonstrate that the framework fits, but to show what each property reveals that the accuracy debate cannot.

## **I. Embeddedness: Why Category Revision Alone Will Fail**

IDEA categories do not exist as freestanding diagnostic constructs. They are woven into the institutional fabric of every school district in the country, determining which students trigger the IEP process. That process generates its own cascade: the referral form, the evaluation, the eligibility meeting, the placement, the staffing assignment, the room. They determine funding flows, because states and districts receive per-pupil allocations tied to disability category, creating financial architecture around each classification. They determine legal obligations, because once a student is classified under IDEA, a district's duties toward that student are categorically different from its duties toward the same student without a label. They structure teacher preparation programs, organized by disability category, mild/moderate versus severe/profound, for generations. They structure the assessment batteries used in eligibility determinations, which are designed to produce evidence that IDEA's categorical structure can process.

*This embeddedness generates a prediction that accuracy analysis cannot: reform targeting category definitions alone will fail.*

The accuracy debate tends to produce recommendations like “improve assessment validity,” “reduce evaluator bias,” or “revise eligibility criteria.” These are reasonable recommendations. They will produce limited sustained effects. Not because they’re wrong, but because they address one node of a system where every node is connected to every other.

The 2004 IDEA reauthorization offers evidence. It introduced Response to Intervention as an alternative identification pathway for Specific Learning Disability, a deliberate attempt to replace the validity-challenged ability-achievement discrepancy model with something more empirically grounded. A decade later, RTI implementation remained highly variable, and SLD identification patterns had not substantially changed (Fletcher et al., 2019). The criteria changed. The infrastructure didn’t. The category is a bureaucratic necessity long before it is a scientific one. Bureaucratic necessities are governed by switching costs, not by scientific validity.

## **II. Installed Base: Disparities as Inheritances**

Current IDEA categories were not designed from scratch in 1975. They were built on top of existing state classification systems, which were built on top of medical and psychological models of disability, which were built on top of the institutional history of American special education.

In 1968, the pediatrician Lloyd Dunn published a paper that launched the disproportionality conversation with a direct finding: disability categories were being used to maintain separate educational programs for students of color after formal racial segregation became legally impermissible (Dunn, 1968). Dunn, writing in *Exceptional Children*, was direct: the special education classes for “mildly retarded” students he was examining were disproportionately filled with Black and Latino children, the education they received was inferior to what they would have received in general education, and the labels were doing work the word “colored” had done before.

IDEA was, among other things, a response to this critique. It mandated procedural protections, required placement in the least restrictive environment, and built in monitoring requirements. It did not rebuild the classification infrastructure from the ground up. It plugged new requirements into an existing system. That system carried its history forward.

*Racial disparities in special education classification are not aberrations to be corrected. They are inheritances.*

Infrastructure theory calls this path dependency: the observation that each version of a classification system inherits the constraints of whatever it was built on, and that each version generates new research, new legal precedent, new professional identities, and new institutional arrangements that constrain the next version. The 2004 reauthorization could not abandon the categorical structure because fifty years of special education practice had been built around it: the rooms, the staff, the funding formulas, the IEPs.

The current system was built on top of a racially structured prior system, and path-dependent infrastructure carries its history forward whether or not its current operators intend it to. This is not an argument that nothing can change. It is an argument that change requires attending to what the system was built on, not only to how its current operators are performing.

The empirical record is consistent with this reframing in a specific way. Singer, Palfrey, Butler, and Walker (1989) found that identification rates for learning disability varied by a factor of nearly four across school districts within a single state, variation that cannot be explained by corresponding variation in the actual prevalence of learning disability. The children in high-identifying districts do not have nearly four times the rate of LD as children in low-identifying districts. What varies is the local classification infrastructure: the referral thresholds, the assessment cultures, the institutional norms about who gets referred, the district's history of how categories have been used.

More recent work by Stiefel, Fatima, Cimpian, and O'Hagan (2024) confirms that school context accounts for substantial variance in identification rates that student-level characteristics cannot explain. The rates are outputs of the infrastructure. The accuracy debate is trying to measure a signal that the infrastructure is generating.

### **III. Visibility Upon Breakdown: The Monitoring Mechanism**

When the classification infrastructure works (student referred, evaluated, placed, served), it disappears. The school psychologist doesn't experience herself as applying a historically contingent, politically constructed classification system. She experiences herself as identifying a child who needs help. This transparency is infrastructure operating exactly as intended. It is also, as Star observed, when infrastructure is most consequential: when the tool becomes invisible, its assumptions become invisible too.

Infrastructure becomes visible upon breakdown. A parent whose child has been placed in a self-contained behavioral classroom asks the school for the data on how often Black students are identified with Emotional Disturbance compared to white students, and the district doesn't have it in a form it can share. An advocacy organization tries to determine whether a state has quietly lowered its risk ratio thresholds to reduce the number of districts flagged for disproportionality, and cannot, because no federal repository tracks methodology changes across states. The system becomes visible when someone pulls at a thread and finds that no one was watching.

The Significant Disproportionality provisions of IDEA, requiring states to calculate and report risk ratios, to flag districts where racial disparities exceed state-defined thresholds, and to direct resources toward addressing them, are the mechanism by which the classification infrastructure becomes visible to anyone outside the institution conducting the classification.

*The monitoring data doesn't fix the infrastructure. It makes the infrastructure legible.*

This distinction is, in Star's framework, structural rather than rhetorical. And it is this distinction, between repairing the system and making the system visible, that the 2025 federal proposal collapses.

## **The Looping Problem**

Infrastructure analysis reveals the structure of the system. It does not specify what happens to the human beings the system processes. The philosopher Ian Hacking spent years working on exactly that question, and his answer is the necessary supplement to Star's framework (Hacking, 1995).

Hacking argued that human classification differs from natural classification in a fundamental way: the classified objects respond to being classified. Gold does not alter its atomic structure because chemists classify it as element 79. A child classified as emotionally disturbed does not have the same non-response. She is placed in a behavioral support classroom: surrounded by other students classified similarly, supervised by staff trained in behavior management rather than academic instruction, assessed through behavioral lenses in every subsequent evaluation, tracked into disciplinary intervention systems that monitor and document the behavior the classification predicted. The classification creates the institutional environment. The institutional environment produces the behavioral patterns. The behavioral patterns confirm the classification.

*This is not a failure of the system. It is the system working.*

And looping effects operating inside infrastructure are more consequential than looping effects in clinical settings because the institutional cascade is harder to interrupt. A skilled clinician can revise a diagnosis when a patient's presentation changes. The room, the track, the legal record triggered by an IDEA eligibility determination do not revise themselves. The professionals who encounter this student in year three of his IEP have been trained to work within the framework the classification established, not to question whether the framework was right.

Copeland, Shanahan, Costello, and Angold (2013) found that most children who receive a psychiatric diagnosis in childhood do not retain the same diagnosis into adulthood. The classification system treats categories as properties of the child. The longitudinal data suggests they are properties of the interaction between the child, the category, and the institutional environment the category creates.

## **The Check Engine Light**

In 2025, the federal government proposed eliminating the Significant Disproportionality data collection requirement under IDEA. The stated rationale was regulatory burden reduction. The

estimated cost savings of the data collection across all states and territories: \$26,880. Total. Across all states. For the only mechanism that makes the classification infrastructure visible.

The Council of Parent Attorneys and Advocates put it plainly: eliminating this requirement would be the equivalent of removing the check engine light from every car on the road and calling it a safety improvement because the light was annoying (COPAA, 2025).

Below the dollar figure, a second structural move. The Office of Special Education and Rehabilitative Services, the federal office responsible for enforcing IDEA compliance, was simultaneously targeted for staffing reductions. Remove the visibility mechanism. Defund the enforcement capacity.

*This is not a paperwork decision. It is a structural choice to make the sorting machine invisible again.*

## What Would It Take

The nine-year-old is still in the room. He will be in the room tomorrow, and next year, and for the rest of his K-12 education. The classification has activated. The institutional environment is reshaping itself around the label. The loop is running. Whatever this essay argues about infrastructure and installed bases and path dependency, it argues about him. So the question is not academic: what would it actually take to change the machine he is inside?

Thirty years of single-node reform have not changed the outputs. The accuracy debate keeps producing the same recommendations: improve assessment validity, reduce evaluator bias, revise eligibility criteria, train evaluators in culturally responsive practice (Waitoller & Artiles, 2013). These are not wrong. The people doing that work are doing real work under real constraints. But the embeddedness analysis predicts, and the 2004 IDEA reauthorization confirms, that single-node reforms get absorbed by the infrastructure without altering its outputs. The criteria changed. The sorting machine did not.

*If the machine is the problem, changing the machine requires working on more than one part of it at once. Five layers. In order. Each one depends on the one beneath it.*

**First: do not let them turn off the lights.** The Significant Disproportionality data collection is the only mechanism that makes the sorting machine visible to anyone outside it. The federal government has proposed eliminating it to save \$26,880. That is not a paperwork decision. That is a decision to make the machine invisible again. Protecting and expanding the monitoring data is the non-negotiable floor. Without it, nothing else on this list is possible, because nobody outside the institution can see what the institution is doing. But visibility alone is not enough. The data has

existed for years. It has not, by itself, transformed identification patterns. The light has to be on before you can see what needs fixing. Turning on the light is not the same as fixing it.

**Second: get the children out of the loop.** Right now, a child who is classified enters an institutional environment that confirms the classification. The loop runs. Nobody interrupts it. Interrupting it requires mechanisms most districts do not have: mandatory reclassification reviews on fixed timelines, conducted by teams that include at least one person who was not part of the original determination. It requires genuine enforcement of least restrictive environment, not the compliance theater version where the paperwork says LRE and the child spends four periods a day in a self-contained classroom (McLaughlin, 2010). And it requires separating the delivery of services from the persistence of the label, so that a child can get what she needs without the classification following her into every subsequent evaluation, every subsequent school, every subsequent year.

**Third: break the link between the label and the money.** The current funding architecture is the engine that keeps the sorting machine running. States and districts receive per-pupil allocations tied to disability category. The label is a fiscal event before it is an educational one. Every incentive in the system points toward classification, because classification is where the money is. Decoupling funding from categorical labels, moving to census-based or functional-need allocation, removes that incentive (Vermont Agency of Education, 2018). Vermont has been doing this since the 1990s: special education funding based on total enrollment, not disability headcount. It works. It does not require replacing the categories. It loosens their grip on the funding node, which is connected to everything else.

**Fourth: stop auditing the outputs. Start auditing the machine.** This is the intervention the accuracy debate cannot see, and the one this essay exists to propose. Disproportionality monitoring currently audits rates: are the outputs of the classification system distributed equitably? An infrastructure audit would examine the system itself: the referral thresholds, the assessment culture, the local history of how categories have been used, the district's inherited practices, the composition of evaluation teams, the availability of alternatives to classification (Aron & Loprest, 2012). This is what Singer et al.'s 1989 finding of fourfold variation across districts is actually measuring. The variation is not in the children. It is in the local classification infrastructure. Nobody is currently proposing this. Someone should.

**Fifth: name the horizon.** The long-term goal is non-categorical service delivery. Children receive educational support based on what they need, not what label they carry. New Zealand's functional needs-based model is the clearest international example (New Zealand Ministry of Education, 2022). In the American context, fifty years of law, practice, professional training, and institutional identity stand between here and there. The switching costs are enormous. But naming the direction matters. It clarifies where every intermediate reform should be pointing. And it makes visible the thing the current system makes invisible: that there are other ways to do this. The sorting

machine is not the only possible machine.

*The accuracy debate asks: are the rates right? This essay asks: what would it take to change the machine? The answer is not one thing. It is five things, in order. And the first one is under active threat right now.*

Will all five layers work as predicted? The honest answer is that this ordering is generated by a framework, not validated by implementation. Decoupled funding might produce under-identification. Infrastructure audits might become compliance theater. Reclassification reviews might be absorbed by the same institutional culture they are meant to interrupt. Every layer carries risks. But thirty years of single-layer intervention have not changed the outputs of the sorting machine. The question is not whether a multi-layer approach carries risk. The question is whether we are willing to keep doing what has not worked and calling it reform.

## The Return

Return to the room. The school psychologist, the teachers, the assistant principal, the mother who took a half day. The nine-year-old. The infrastructure worked exactly as designed. The question was never whether the people in that room performed well. They did. The question is what kind of system put them there, what it was built on, and what it costs to see.

*Every classification system has a designer.*

*Every design has a politics.*

*Every installed base has a history.*

The sorting machine is not sorting randomly, and it is not sorting optimally. Understanding what kind of machine it is, rather than only measuring whether it is measuring correctly, is the analytical task this essay opens.

---

## References

- Aron, L., & Loprest, P. (2012). Disability and the education system. *The Future of Children*, 22(1), 97–122.
- Blanchett, W. J. (2006). Disproportionate representation of African American students in special education: Acknowledging the role of white privilege and racism. *Educational Researcher*, 35(6), 24–28.
- Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press.
- Copeland, W. E., Shanahan, L., Costello, E. J., & Angold, A. (2013). Cumulative prevalence of psychiatric disorders by young adulthood. *JAMA Psychiatry*, 68(10), 1023–1030.

- Council of Parent Attorneys and Advocates (COPAA). (2025). Comment letter on proposed elimination of Significant Disproportionality data collection.
- Dunn, L. M. (1968). Special education for the mildly retarded: Is much of it justifiable? *Exceptional Children*, 35(1), 5–22.
- Fish, R., Shores, K., & Souto Maior, J. (2026). Covariate adjustment and the construction of disability in disproportionality research. *Exceptional Children*.
- Fletcher, J. M., et al. (2019). Response to intervention and the identification of specific learning disabilities: A review. *Learning Disability Quarterly*, 42(4), 233–244.
- Hacking, I. (1995). The looping effects of human kinds. In D. Sperber et al. (Eds.), *Causal Cognition: A Multi-disciplinary Debate* (pp. 351–394). Oxford University Press.
- McLaughlin, M. J. (2010). Evolving interpretations of educational equity and students with disabilities. *Exceptional Children*, 76(3), 265–278.
- Morgan, P. L., et al. (2015). Which racial/ethnic groups are most affected by special education identification? *Educational Researcher*, 44(5), 278–292.
- New Zealand Ministry of Education. (2022). Learning support delivery model: Functional needs-based allocation framework. Wellington, NZ.
- Singer, J. D., Palfrey, J. S., Butler, J. A., & Walker, D. K. (1989). Variation in special education classification across school districts. *American Educational Research Journal*, 26(2), 261–281.
- Skiba, R. J., et al. (2016). Risks and consequences of oversimplifying educational inequities: A response to Morgan et al. (2015). *Educational Researcher*, 45(3), 221–225.
- Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43(3), 377–391.
- Stiefel, L., Fatima, Z., Cimpian, J. R., & O'Hagan, E. (2024). School context and special education identification rates. *Journal of Education Finance*.
- Vermont Agency of Education. (2018). Non-categorical special education funding model: Implementation report. Montpelier, VT.
- Waitoller, F. R., & Artiles, A. J. (2013). A decade of professional development research for inclusive education: A critical review and notes for a research program. *Review of Educational Research*, 83(3), 319–356.