



When Learning and Assessment Diverge: Who and What are We Assessing?

James Paul Gee
Arizona State University

Both our assessments and tests and our research on assessments and tests have not kept pace with ongoing modern research on the human mind and human learning. Too often we do not know *who* we are assessing when we assess students today. A student is not just a data point. Rather a student is a complex body of experience gained over (long periods of) time, stored in mind in different ways, and recruited as a source of hypotheses, meaning making, and planning when the student needs to act, learn, or take a test. If we want to render assessments fair, useful, and meaningful—if we want to be able to act intelligently on their outcomes—we need to see that we should be assessing bodies of experience and judging how they can be used now and in the future for better learning and development and what new experiences the student needs for future growth and the building of new capacities.

Many standardized tests are what I will call “target assessments”. Target assessments tell us how many people out of a designated population have hit a certain target, for example how many can compute fractions at a certain level of correctness.

Some target assessments just tell us how many people can hit the target and leave it at that. Let’s call these “cognitively superficial target assessments”. Other target assessments tell us how many people not only can hit the target, but understand what the target (e.g., fractions) means and why it works as it does. Let’s call these sorts of assessments “cognitively deep target assessments”.

Some research indicates that cognitively deep target assessments correlate with the ability to use what one knows to solve relevant sorts of problems in the world, while cognitively superficial ones do not. Of course, we can design target assessments that assess how many people can actually use what they know in a relevant domain—say, fractions—to solve relevant real-world

problems to whose solutions fractions contribute. We can call these “real-world problem solving target assessments”.

Target assessments on large enough samples can, perhaps, have policy uses, but they tell us very little about human beings and next to nothing about individuals. The reason this is so is that when we assess an individual we need to know not how good that person is at hitting a target, but how good that person is *relative* to how much preparation he or she has had in terms of target practice. Obviously a newcomer that gets fairly near the bull’s eye on a target is better than the veteran who consistently misses by the same amount. It would quite unfair and quite uninformative for most purposes to hold the newcomer to the same standards in terms of outcomes as the expert (although we do want to hold both to the same standards as a goal or aspiration). And we should remember that thanks to the inequities in our society and schools very often—even at the college level—the same class or cohort contains students far apart on the continua of newbie to expert based on past experience, prior practice, and access to relevant preparation for learning.

Modern research on situated and embodied cognition gives us some strong indicators about what constitutes good “target practice” for human beings. An assessment is only fair and informative for most purposes when we are assessing a target relative to what I will call the “inputs” to the target and these inputs are the various types of good target practice a person has had

Here are some features of what constitutes good target practice for humans and, thus, what needs to constitute inputs to relative judgements about hitting targets. First, the amount of previous situated (situational) experience and practice a person has had in a domain (e.g., geometry, gardening, gaming, and composition) is a crucial input to performance in the domain. For example experiences with seeing how algebraic equations actually apply to specific situations is an important input. This is so because, according to work on situated and embodied cognition, in many cases, humans understand abstract things (like equations) initially bottom up from concrete applications. So they need to have had ample concrete experience, not just abstractions, symbols, texts, and words. Let us call this sort of situated experiential target practice, “situated experiential input” (“i^{se}” for short).

Second, some experiences are better than others for learning. Just being left on one's own is not as helpful as receiving good mentoring, at least before a person has become adept enough at learning in a domain to be a deliberate learner (self-teacher). When people get well-designed, well-mentored experiences where they have a clear goal, have an action to take whose outcome they care about, and are helped to manage their attention well (in order to avoid cognitive overload) they learn much better and are much better prepared for future learning experiences. This is so, because they integrate the relevant aspects of their experience better with prior knowledge and, thus, can use these elements as good preparation for future learning. So it is important to know not just how much situated experience a person has had but how much of this situated experience has been well-designed and well-mentored. Let us call this sort of target practice “well-designed experience input” (“i^{wde}” for short).

Third, people learn well only when they are motivated and care about what they are learning (which, for humans, means that something is felt to be at stake or matter for the person in regard to what is to be learned). This often amounts to whether the person “affiliates” with what is to be learned and it is especially important that the person does not actively disaffiliate with it, since then an affective filter is present and blocks input from being in-take (taken in and actually processed). So we need to know how much experience a person has had where they have affiliated with the goals and values of what is to be learned and surely not actively disaffiliated from them. Even if they have had lots of experience, if their affective filter has been up this experience has to be discounted (in either the sense of ignored or in the sense of less valued). Let's call this “affiliated experience input” (“i^{af}”).

People over time (and this often requires a significant amount of time) use concrete experience to find patterns and sub-patterns. Over time, as they have more and more relevant experiences in a domain, they come to find and be able to use more and more general patterns and their knowledge becomes more abstract. This is part of why the amount of relevant experience a person has had matters: people need time (and teaching and mentoring) to find useful patterns and sub-patterns and to generalize them more and more. So we need to know how far along a person is in pattern-recognition (generalization)—how much experience they have had at the

point where how they process new experiences is becoming more general, abstract, meta-cognitive, and strategic. Let us call this “meta experience input” (“i^m”).

Too often in schools, we take reading texts as the main form of experience necessary for learning. But humans give situational (contextually) appropriate meanings to the words in texts based on experiences they have had in the world, in media, in talk and dialogue, and in reading other texts. Words and texts give categorical meanings to the world, but the world gives contextually appropriate and nuanced meanings to words and texts. So we need to know a good deal more than what a person has read in order to make relative (fair) assessments about target hitting.

So, there are number of different types of input (i) all of which are relevant. A fair and informative assessment would assess how good a person is at hitting a target relative to how much input the person has had in terms of all these different types, that is t/i^d where “t” is a target and “i^d” is the distribution of the amounts of the various types of input we have just discussed. Let us call such assessments—in most cases, aside from some relatively abstract policy issues, the only types that are fair and informative—“trajectory target assessments” (TT assessments), since we are assessing a target relative to a person’s trajectory through experiences that lead to being able to hit the target more and more.

While it would be true in any country that TT assessments would, for the most part, be the only fair and informative ones, this is particularly true in the United States where access to the various forms of input we have discussed in regard to certain sorts of targets is highly unequally distributed based on race, class, sometimes gender, and often in regard to people who are “different” (which, in actuality, means most of us in some domains).

Note that the issue here is not just developmental stages. In a trajectory approach we are focused on the fact that in a great many domains (e.g., algebra, film making, gaming, cooking, medicine) there are a series of gates along one or more paths to mastery that must be passed along the way, and time on task (in experiences) is required between each gate. These gates do bear a complex relation to development (and, in some cases, perhaps, stages), but they are partly inherently

connected to the internal “logic” and external social organization (in terms of practices) of the domain. Furthermore, in many domains there are different trajectories to mastery with somewhat different gates placed in somewhat different places along the way.

Imagine a race in which some people got to run ten minutes and other people got to run three minutes and then we judged the winner by who went the furthest. This would be manifestly unfair. What we would need to do is judge how far each person went relative to the time they had to run. Even this analogy does not capture the full amount of unfairness here though, since in most important domains in life it does not matter how fast one gains mastery (it really is not a race), just that they do, and sometimes people who take longer gain better or more innovative forms of mastery.

So if it is manifestly unintelligent to assess people other than on TT assessments, why do we do so and do so all the time? There are, of course, institutional reasons having to do with practicality and standardization (and profit), but another reason is that we have a wrong-headed view of opportunity to learn. We think opportunity to learn means that people have had access to the same books and been in the same grades. But it really means that people have had equally many rich previous experiences (inputs) relative to targets in the domain being assessed.

TT assessments would tell us what a person can do relative to what they have done in the past and where they are in a trajectory towards mastery. And this is useful. But they cannot tell us what people COULD do if either they had more and better inputs or were to be set off on a new trajectory in a domain they did not know about or previously had no access to. TT assessments—while fair and informative—are not as useful (and moral) as they could be if they were not just about the past, but about potential and growing new capacities, as well.

We need TT assessments that tell us about people’s capacities for growth and how these capacities might fit with their talents, interests, and passions. This is an absolutely crucial issue in a society that so ill-distributes possibilities for certain sorts of experiences and access to certain sorts of domains. It is pretty useless to tell me that someone is a bad writer when they have had little experience with writing, even sillier when they have had none (even if they have

sat in writing classrooms). In this case, what I would want to know is what are this person's potentials here; how can he or she best be resourced; and how do these resources fit with or help create this person's talents, interests, passions, and values as a unique individual (not just a member of a certain race, gender, class, or disability group)?

So we need assessments that discover, resource, and build capacities. These are assessments *for learning* not of learning, but they are more than that, they are also assessments for *futures*. In such assessments we would need to lay out a path (trajectory) of gates and ensure that the person gets good and copious (good) inputs towards each gate and then see how they fare, both in terms of performance, but also in terms of interests and passions, because, after all, there are other paths and the important thing is to get on good ones for one's own unique contribution to society. Let's call such assessments "preparation for future targets assessments" (PFT assessments).

TT assessments (assessments that are relativized to a given distribution of inputs) and PFT assessments (assessments that are based on new inputs of the right sorts) are not "mainstream". This is so, in part, of course, because such assessments are much harder to standardize, use at scale, and render psychometrically tractable than current standardized tests, at least. But we surely do not want efficiency and psychometric elegance to trump useful information, fairness, morality, and impact for good on society and the world. An efficient and tractable assessment can still be useless or even a lie.

In my view, we need to constitute a wider sort of field of assessment—one with contributions from more disciplines than psychometrics alone, but all integrated in a powerful way—to engage in research and interventions that speak directly to TT assessments and PDT assessments. I believe this would change "the game" in several different senses and for the better in a variety of different ways. The point here is that as ideas about learning and assessment/testing—and the sorts of data new information technologies can gather, mine, and analyze—diverge further and further we face deep problems that are both practical and moral.

It is also the case that modern work on memory and how we assess and test are also diverging. Modern research shows that human long-term memory is a storage space of edited experiences

people have had with some degree of organization into patterns and sub-patterns across experiences. This long-term memory is oriented more to the future than the past in the sense that human memory is not very good at accurate recall, but is primarily oriented to be used as fodder for planning, hypothesizing, acting, reflecting, and learning in the future when one needs to solve problems. Our assessments are often more oriented to the past—to recall—than to the future—capacity building, potential, and growth. What counts for humans in human memory is where the person is going, not just where they have been. This is so for good Darwinian reasons: we survive by using the past to move to the future, not by lingering in the past. It is time, I think, for assessment and testing to move into the future, as well.