ACT's Efficacy Framework:

The Intersection of Learning, Measurement, and Navigation

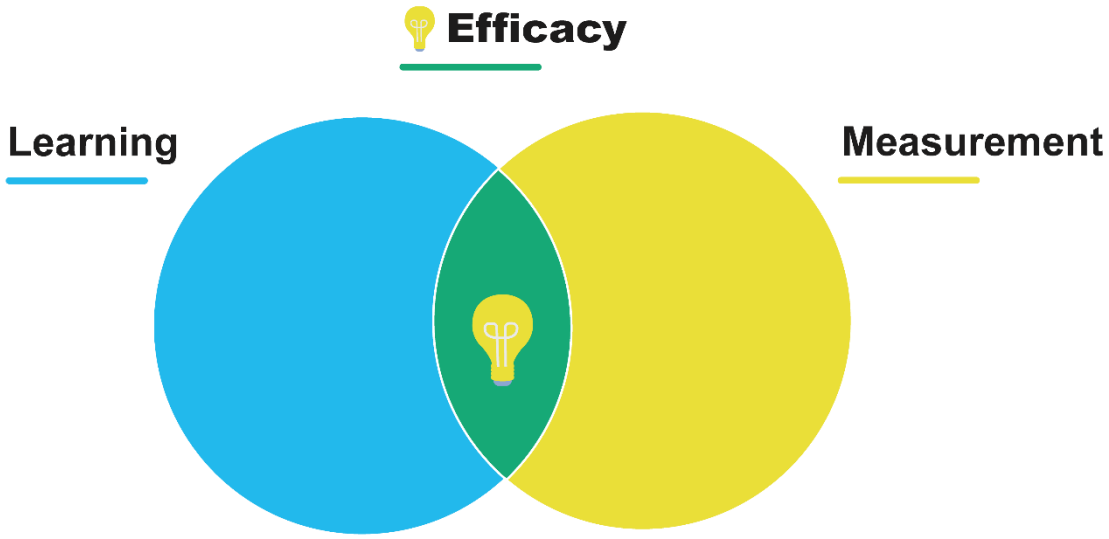
Krista Mattern, PhD

Abstract

A great deal has been written on the topic of test validity. Guiding our work at ACT are *The Standards for Educational and Psychological Testing* (2014), which outlines best practices in test development and validation. As ACT transitions from an assessment company to a learning, measurement, and navigation organization, a framework for our learning products is also needed to guide research activities to ensure that the hallmarks of rigor, science, and scrutiny of our measurement solutions are also built into our learning products. The field of education and education measurement has devoted much less attention to the development of efficacy arguments and frameworks as compared to validity arguments and frameworks. More importantly, the integration of validity theory and efficacy theory into an overarching Efficacy Framework is needed to ensure that we optimize the conditions to stand up an evaluation system with appropriate feedback loops so that we can appropriately assess whether our solutions are designed to have the greatest impact on learner outcomes. This point is exceedingly critical because it lies at the heart of ACT's mission. If we cannot evaluate whether our products are efficacious, then how do we know if we are achieving our mission of helping individuals achieve educational and workplace success? The Efficacy Framework presented here is an attempt to guide research activities at ACT in support of our mission.

In this paper, we advance the proposition that the quantification of efficacy is achieved via the intersection of learning and measurement, as visually displayed in Figure 1. In order to evaluate whether a learning tool is efficacious, we need to measure the impact of the learning tool on the intended learner outcome.

Figure 1. ACT's Efficacy Framework: Intersection of Learning and Measurement



Validity Argument for ACT Measurement Solutions

As described in *The Standards,* **validity** refers to the degree to which evidence and theory support the interpretation of test scores for proposed uses of tests (2014). In reference to efficacy, we need to collect evidence to evaluate the degree to which test scores support inferences of student learning. For example, the appropriateness of using a specific test to detect whether the learning experience changed the targeted knowledge and skills in the intended manner needs to be evaluated, highlighting the importance of coherence across learning tools and assessments.

Types of Evidence Supporting a Validity Argument

The Standards identify five sources of validity evidence for tests. A description of each of the sources of validity evidence is provided below.

1. *Evidence Based on Test Content* – The degree to which the content of the test reflects the construct(s) it intends to measure.

2. *Evidence Based on Response Processes* – The degree to which assumptions about the cognitive processes engaged by test users occur.

3. *Evidence Based on Internal Structure* – The degree to which the relationships among measurement opportunities and test components conform to the construct on which the proposed score interpretations are based.

4. *Evidence Based on Relations to Other Variables* – The degree to which relationships with other variables are consistent with the construct underlying the proposed score interpretations.

5. *Evidence Based on Consequences of Testing* – The degree to which the expected benefit from the intended use of test scores are realized.

Efficacy Argument for ACT Learning Solutions

***Efficacy* refers to the degree to which evidence, rationales, and theory support the claim that a learning tool improves intended learner outcomes under ideal conditions.**

It is important to note that the literature often makes a distinction between efficacy and effectiveness where efficacy of an intervention is its optimal effect realized under ideal conditions whereas effectiveness is an intervention's typical effects realized under normal conditions. Along those lines, ACT's Efficacy Framework includes Use and Implementation Fidelity as a source of efficacy evidence to promote the design and execution of research studies that not only examine whether a learning tool is effective but to identify conditions and implementation models that optimize learning.

Outcomes Framework

Before evaluating the efficacy of a learner tool, the outcome(s) (e.g., college remediation rate) we intend to impact need to be specified. The outcome of interest can also be framed in terms of a customer problem that we hope the learning product will help solve. For example, higher education institutions may have a high percentage of students that need to take remedial course work and wish to lower that percentage. Kirkpatrick's training evaluation model (1959, 1976) is a popular model in industrial-organizational psychology for evaluating the impact of job training programs on workplace outcomes. Other researchers have highlighted the utility of applying this model to the education space, such as the evaluation of higher education learning outcomes (Praslova, 2010). Given the simplicity and flexibility of the

model, we contend that the model can be applied to any learning outcome regardless of the population (students, workers) and setting (school, work). As such, we propose adapting Kirkpatrick's model (Figure 2) to research the efficacy of learning products at ACT. The model classifies learner outcomes into four categories beginning with more proximal or immediate outcomes located at the base of the pyramid to more distal outcomes located at the apex of the pyramid:

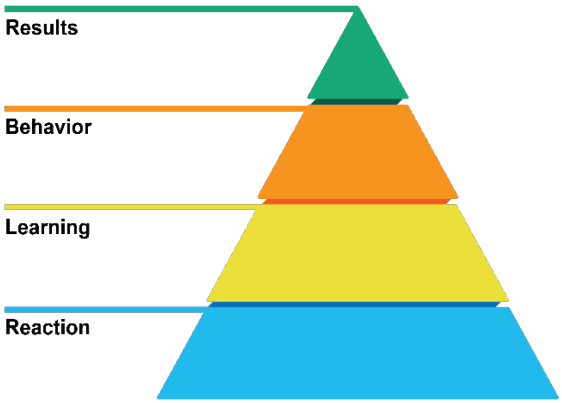
1. *Reaction* – The degree to which individuals find the learning event favorable, engaging, and relevant

2. *Learning* – The degree to which individuals acquire the *intended* knowledge, skills, abilities, and other characteristics (KSAOs) based on their participation in the learning event

3. *Behavior* – The degree to which individuals apply what they learned from using the learning tool to real-life events (classroom performance, job performance)

4. *Results* – The degree to which intended outcomes occur as a result of using the learning tool. For example, the degree to which increased learning results in learner outcomes we wish to positively influence such as higher college enrollment, lower remediation rates, and persistence and graduation rates.

Figure 2. Kirkpatrick's Evaluation Model



Types of Evidence Supporting an Efficacy Argument

The Standards (2014) outline five sources of validity evidence. We adapt those sources of evidence to be applicable to learning products as well as include additional sources of evidence to incorporate Kirkpatrick's training model described above. The resulting seven sources of efficacy evidence are:

1. *Evidence Based on User Experience* - The degree to which individuals find the learning tool favorable, engaging, and relevant.

2. *Evidence Based on Content* - The degree to which content delivered in the learning tool is of high-quality and aligned to the content of targeted outcome(s), such as course curriculum or targeted standards. Content is in reference to what is covered in the learning resource and measured via an assessment; it is not referencing content knowledge. Therefore, alignment of content is not limited to content/declarative knowledge but could also include procedural knowledge, skills, and abilities and/or practices.

3. *Evidence Based on Personalization* - The degree to which the content delivered via the learning tool is appropriate to the individual's current level of KSAOs and adapts as an individual improves their KSAOs within the system. Personalization could also take the form of tailoring content to a student's career and personal interests to promote user engagement.

4. *Evidence Based on Learning* - The degree to which individuals acquire the intended KSAOs based on using the learning tool.

5. *Evidence Based on Use and Implementation Fidelity* – The degree to which the effectiveness of the learning tool is dependent on how it is implemented. What is the most appropriate use of the learning tool? Does the magnitude of learning depend on

the implementation model of the learning tool? Does the magnitude of learning depend on other contextual variables?

6. *Evidence Based on Relations to Other Variables* – Degree to which performance in the learning tool is related to targeted outcome(s).

7. *Evidence of Results/Impact* - The degree to which targeted outcomes occur as a result of using the learning tool. Do individuals who use the learning tool have improved outcomes (intended learner outcomes)?

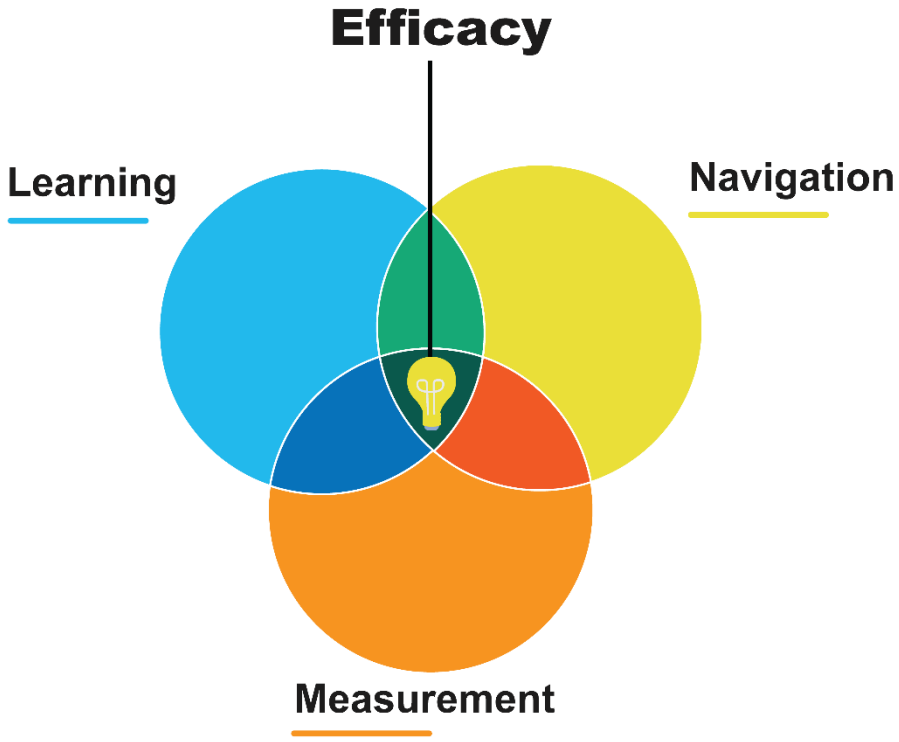
For learning products that make claims about improving learner outcomes that occur far into the future, it is important to test and validate those claims. Given that it will sometimes take months if not years to collect those data, researchers should also plan for short-term efficacy studies where one evaluates the impact of the learning product on intermediate outcomes or potential mediating variables to obtain an early indication of whether or not the product will have a positive impact on these more long-term outcomes. It is important to note that as the time between using the learning product and the collection of results/impact data increases, the likelihood that a small or no effect will be observed increases due to intervening variables, such as subsequent instruction and educational experiences. These additional variables or alternative explanations should be included in the learning product's theory of action (TOA) and incorporated in the design of a longitudinal efficacy study for the purpose of serving as control variables and/or to help contextualize study findings.

Intersection of Learning, Measurement, and Navigation

Whereas ensuring that the content of the learning tool is aligned to the learner outcomes can promote learning, evidence of personalization can amplify that learning. We propose that the overlap of ACT's three strategic pillars – learning, measurement, and navigation – not only constitutes efficacy but represents the amplification of efficacy, as displayed in Figure 3. In particular, adaptive learning – an area that is gaining traction in the online learning space – combines learning and measurement using technology and platforms that bring together the act of learning and teaching, the measurement of that activity, and the criteria framework that defines it. Such a system enables, accelerates, and validates learning. Additionally, learning analytics, which involves the collection and analysis of a learner's holistic data for the purpose of understanding individual or group interactions and activities and optimizing learning experience by leveraging big data, analytics, and empirical research and insights on all aspects of education, also plays a role in this area.

We argue that both adaptive learning and learning analytics support ACT's strategic pillar of navigation, which encompasses the ways that ACT can connect with customers to provide guidance and navigation on their journey through life. Solutions help individuals and their advocates (parents, counselors, employers, etc.) make good, informed decisions. According to this definition of navigation, customers are assumed to be an active participant in the navigation process at more of a macro level. For example, ACT data can help individuals choose appropriate learning products or apply to a good-fitting college or declare a major that is aligned with their interests. However, we can also conceptualize navigation at a more micro and passive level. Through the development of adaptive learning algorithms, we guide or navigate students – perhaps unbeknownst to them – through a sea of learning content and deliver the resources that our research has determined are the most relevant for them and continues to adapt based on their learning, serving up the next most relevant piece of content. Therefore, within the closed system of a learning platform, we ensure students are maximizing their time and experience and thus hopefully their learning by focusing on content they are prepared to master and filtering out content they have already mastered or not yet prepared to master.

Figure 3. Amplifying ACT's Efficacy Framework: Intersection of Learning, Measurement, and Navigation



Development of ACT's Efficacy Framework

The previous section underscored the point that an efficacy argument is dependent on the availability and suitability of an assessment to measure that learning. And the suitability of that measurement tool is dependent on the validity evidence supporting that use of that assessment to estimate learning. As such, ACT's Efficacy Framework requires the integration of both a validity argument and an efficacy argument.

Integration of a Validity and an Efficacy Argument into an Efficacy Framework

Intuitively, ACT already knows learning and measurement must go hand in hand. This is evident by a review of our product portfolio. For each ACT assessment solution, there is also a learning solution counterpart. Specifically, at ACT, we not only provide assessments to measure what individuals know and are able to do but also provide learning solutions to improve those KSAOs if individuals need to skill up or if they desire to further propel their learning. For example, for the ACT®, we have ACT Academy™, AOP, ACT Kaplan® Online Prep Live (AKOPL), and ACT® Recommends™ – all of which are learning tools to help students better prepare for the ACT. PreACT® is also listed as a learning solution. Even though the PreACT is clearly an assessment, it provides students with a realistic preview of their likely

experience on the ACT. In addition, the PreACT score report can provide useful information in terms of areas where students need to improve prior to sitting for the ACT.

Even though implicitly we know the importance of coupling learning and measurement, we do not currently have an Efficacy Framework explicitly describing how to integrate validity and efficacy arguments and evidence into an overarching framework. This is important as we want to accomplish two goals:

1. Develop learning solutions that are most likely to impact the intended outcome.

2. Increase our ability to detect whether a learning solution is achieving its intended outcome through thoughtful study designs, methodology, and data collection.

To this end, we propose an overarching Efficacy Framework, detailing how the alignment between validity evidence of assessments and efficacy evidence of learning products will promote conditions that optimize the likelihood that we will accomplish those two goals.

Alignment of Validity Evidence and Efficacy Evidence

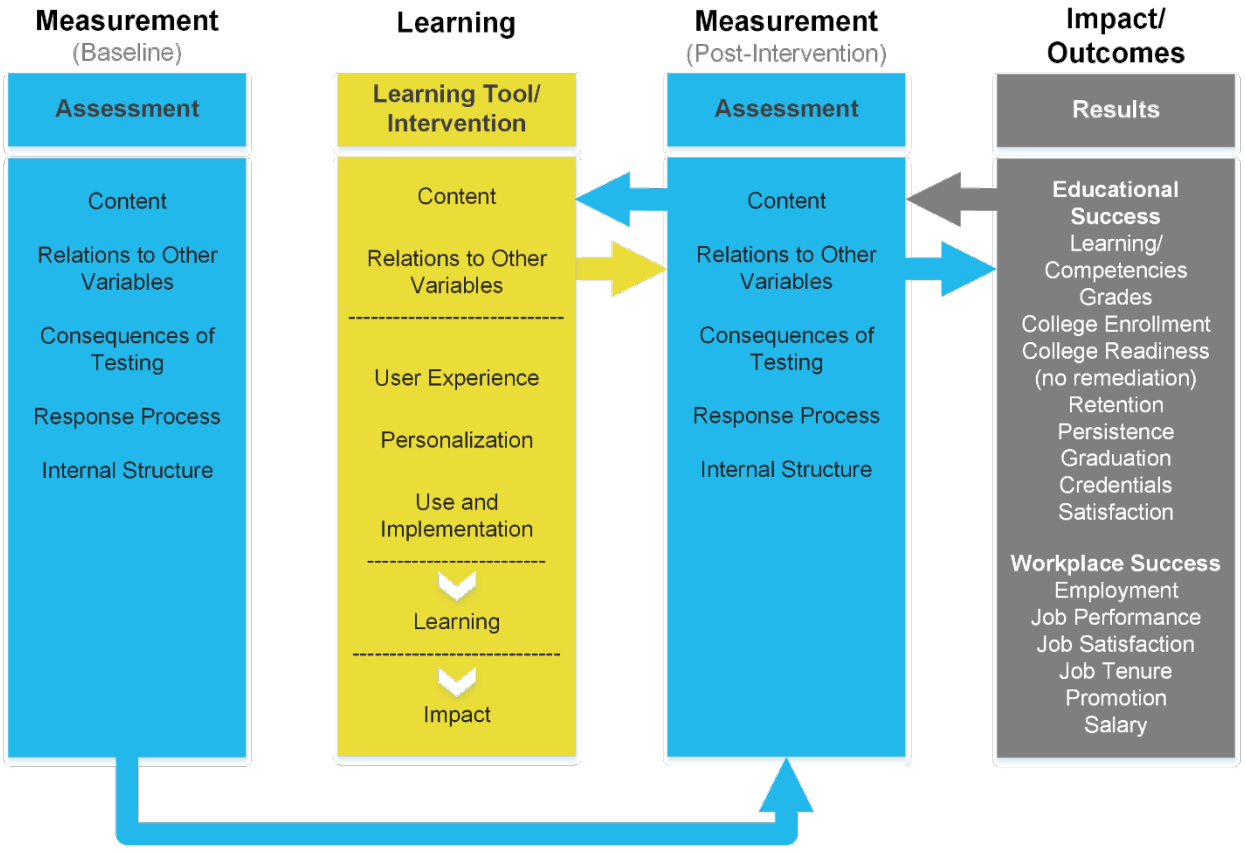
Figure 4 illustrates how validity evidence and efficacy evidence work in concert in the development and evaluation of learning tools. As was described earlier, the first step is to identify the learner outcome that we ultimately want to impact. This is represented in the far right column in the diagram. Examples of learner outcomes we may want to focus on for both the educational and workforce settings are provided. This should not be considered an exhaustive list of possible learner outcomes that ACT is interested in improving. As an example, we may want to increase college readiness among all high school students and thus remove the need for remedial education in college. Once a learner outcome is identified, the next step would be to identify the KSAOs necessary for that outcome – in this example, college readiness. We can focus on the ACT, which is a measure of college readiness, for illustrative purposes. To determine what students need to know and be able to do to succeed in college, ACT routinely surveys college faculty to ask them about the prerequisite knowledge and skills incoming students need to succeed in their content domain. This information guides the ACT test blueprint, which specifies which skills will be included for each subject and how they will be measured. This is represented in the diagram by the grey arrow from the *Impact/Outcomes* column pointing to *Content* of the assessment in the *Measurement (Post-Intervention)* column. To empirically test whether the knowledge and skills assessed on the ACT are the skills that are needed to succeed in college, we conduct research showing that ACT scores are predictive of college success, such as course grades. This is represented in the diagram by the blue arrow from the *Relations to Other Variables* in the *Measurement (Post-Intervention)* column pointing to the *Impact/Outcomes* column.

Now that we have documented *what* students need to know and be able to do to succeed in college, the next step is to develop a solution that helps students improve those exact KSAOs. Therefore, the content covered on the ACT should inform the content that is included in the learning product, which is represented as a blue arrow from the *Measurement (Post-Intervention)* column pointing to *Content* of the *Learning* column. For this example, we will use AOP as the learning solution. To ensure that the content of AOP does in fact align to the content measured on the ACT, we can compare content coverage of AOP to ACT test blueprints (content alignment) or we can examine whether AOP performance is predictive of ACT scores. This is represented in the diagram by the gold arrow from the *Relations to Other Variables* in the *Learning* column pointing to the *Measurement (Post-Intervention)* column.

Within the *Learning* column, there are two downward-pointing arrows, one of which represents how *User Experience*,

Personalization, and *Use and Implementation* can impact the degree to which *Learning* occurs, in addition to the actual learning *Content*. The other represents how the magnitude of *Learning* will influence the magnitude of *Impact*. Finally, the far left column represents the need for measurement of the same KSAOs as the learning solution and post-assessment prior to use of the learning tool to serve as a baseline measure. In some cases, the pre- and post-assessment are the same – in this case, the ACT. Other times, we may propose developmentally relevant, yet still aligned pre- and post-assessments such as PreACT to ACT. This is for similar reasons discussed above indicating the need to link evidence across other columns in the diagram, namely content alignment and predictive validity. Having a baseline measure is useful to estimate individual growth; however, it is not always necessary if students can be randomly assigned to treatment and control conditions and thus should be roughly equivalent on the KSAOs of interest. Even when experimental studies are feasible, it is often helpful to have a pre-measure to check how well the random assignment actually worked.

Figure 4. ACT's Efficacy Framework: Alignment of Validity and Efficacy Evidence



The figure above represents a systems approach to learning and measurement, linking evidence across the two to drive the development of efficacious learning solutions as well as inform future enhancements to the product. The framework also guides the different types of research activities and evidence we should collect. Finally, we propose a framework for determining the level or rigor of evidence needed to guide research activities for specific learning products. In particular, the required level of rigor of evidence should take into consideration several factors when determining the appropriate study design: use case, stage of the product life cycle, desired strength of validity or efficacy claim and funding, intended audience, and return on investment/investment risk.

To read more about ACT's Efficacy Framework, see the full research report available at:
https://www.act.org/content/dam/act/unsecured/documents/R1749-efficacy-framework-2019-06.pdf

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. Washington, DC: AERA.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors, 13*, 3–9.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (2nd ed., pp. 301–319). New York, NY: McGraw-Hill.

Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. *Educational Assessment, Evaluation and Accountability*, *22*(3), 215-225.

Krista Mattern, PhD

Krista Mattern is a senior director in Validity and Efficacy Research whose research focuses on predicting education and workplace success through evaluating the validity and fairness of cognitive and non-cognitive measures. Also known for work in evaluating the efficacy of learning products to help improve intended learner outcomes.