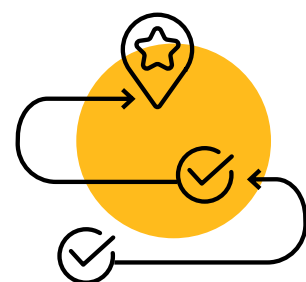Pearson

Learning Design Principles
Purposeful Design

Assessment & Evidence-Centered Design

Summary

What are Pearson's Learning Design Principles?

Our Learning Foundations describe the optimal conditions for learning and reflect the learner experience we hope our products will create. We do this by incorporating our Learning Design Principles.

Each of our Learning Design Principles goes into detail about a key principle, supporting product design and marketing by describing:
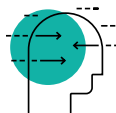
- the research that informs the principle
- why it matters in learning
- how we can apply it in practice

Our portfolio of Learning Design Principles will continue to grow over time.

Welcoming Experience
- Motivation & Mindset
- Social & Collaborative Learning

Minds in Mind
- Developing Understanding
- Attention & Cognitive Load
- Active Learning, Memory & Practice
- Desirable Difficulty & Scaffolding
- Feedback for Learning

Learning Behavior
- Self-Regulated Learning & Metacognition

Purposeful Design
- Objective Design
- Assessment & Evidence-Centered Design
- Personalized Learning & Adaptive Systems
- Authentic Learning

Learn Anywhere
- English Performance Standards
- Digital & Virtual Learning

Assessment & Evidence-Centered Design

For learners to become confident they've mastered the knowledge and skills they need, learning experiences need to be efficient and targeted. Targeted learning experiences rely on accurate assessments of where each learner is.

How do we design accurate assessments that reliably capture these essential insights about learners?

Assessment is any activity that enables us to make inferences about a learner's knowledge, skills or abilities. More than just end of unit tests, assessments can diagnose, support or evaluate learning.

Evidence-centered design (ECD) is a four-component framework for designing assessments to reveal meaningful information about learning and to support learning through effective instructional design.

ECD framework

Domain model

Questions it answers: What knowledge, skills, and abilities are we interested in?

Example: Full list of objectives from a course

Learner model

Questions it answers: What knowledge, skills, and abilities from the domain model do we want to track learners' proficiency in?

Example: A single objective, e.g., 'learners evaluate conflicting historical information'

Task model

Questions it answers: What activities will provide evidence for the learner model?

Example: Tasks aligned to an objective, e.g., reading passage stimulus with open-ended response; scenario with selected-response items; persuasive essay prompt

Evidence model

Questions it answers: What specific behaviors can we observe in the activities that inform the learner model?

Example: Learner behaviors resulting from a task, e.g., recognition of and response to conflicting evidence

Why it matters

Well-designed assessments can:

- support learning by providing accurate, rich and effective feedback and insight on status and progress
- strengthen the accuracy of claims made about learning status and progress
- support clear ways to measure learner outcomes
- show mastery or achievement in a valid and reliable way

Evidence-centered design is a framework for creating assessments that:

- enable valid inferences about knowledge, skills, and abilities
- go beyond traditional assessment activities (e.g., simulations instead of multiple choice)
- can cope with complex and innovative measurement needs
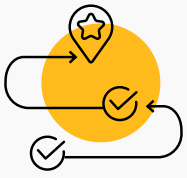
The big ideas

1

Assessments can diagnose, support, and evaluate learning. It's essential to design assessment **with the purpose in mind** to make sure assessment provides valid evidence.

Assessments let me check my knowledge level. They also give me feedback while I'm practicing, and I trust them to help me figure out whether I'm on track.

I can tell that the specific thing I'm about to learn is a part of the wider field I want to get better at.

2

The starting point for ECD is the domain model — a clear **lay of the land** of the learning domain that helps plan effective instruction and gives an accurate idea of a learner's knowledge level.

3

Showing and tracking the development of understanding in a domain, or **learner proficiency**, over time is needed to target instruction design and personalized learning. This is the learner model.

I can track where I'm proficient and what I still need to learn.

By doing this task, I can find out if my learning is on track.

4

Learning tasks or activities reveal evidence about what a learner knows and can do. Tasks should be **designed to capture evidence** of success, or learner behavior that reflects the desired objective or outcome. This is the task model.

5

All the ways a learner could respond to a task count as evidence and provide **insights into their proficiency** and understanding. This is the evidence model.

The way I respond to this task provides insights into my understanding and abilities.

I can trust the assessment to measure what I know and what I need to learn.

6

Assessments should consistently and accurately **measure what they claim to**. A reliable, valid, and fair assessment means we can make sound claims about learning.
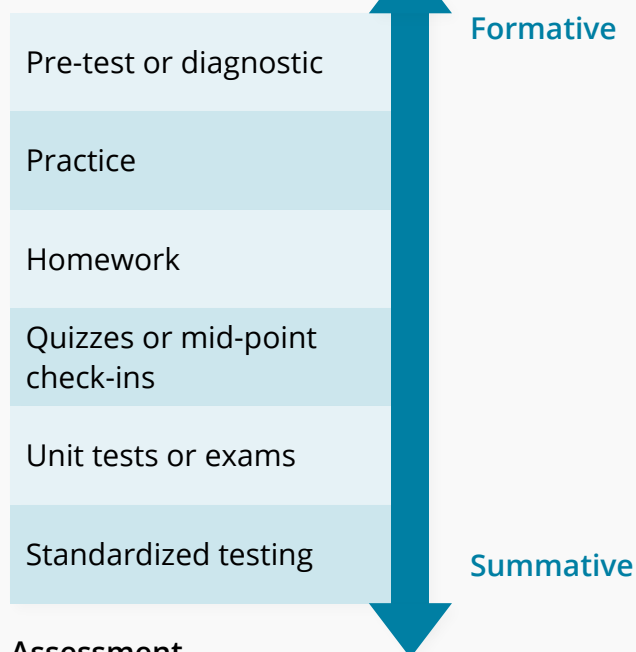
Design for purpose

Assessments can diagnose, support, and evaluate learning. It's essential to design assessment with the purpose in mind to make sure assessment provides valid evidence.

What it feels like for learners

Assessments let me check my knowledge level. They also give me feedback while I'm practicing, and I trust them to help me figure out whether I'm on track.

Assessment for learning

Pre-test or diagnostic
Practice
Homework
Quizzes or mid-point check-ins
Unit tests or exams
Standardized testing

Formative

Summative

Assessment for achievement

- **Diagnostic assessment** evaluates a learner's proficiency before learning and pinpoints areas for instruction. It points towards corrective interventions or indicates the learner is ready to move on.
  - **Purpose:** To find out where a learner is before starting a new learning experience
- **Formative assessment** provides opportunities for learners (and instructors) to gauge progress, receive feedback, and adjust accordingly.
  - **Purpose:** To support learning during instruction
- **Summative assessment** is a chance for the learner to demonstrate that they know what they're doing by applying knowledge or skills
  - **Purpose:** To determine mastery or achievement
- Formative and summative assessments are designed the same way, but serve different purposes

What it means for designing learning experiences

- Make sure assessment types are based on the purpose of the assessment and the nature of the learning objectives and skills — diagnostic, formative, summative, performance, self, etc.

All assessments should include:

- domain model with objective hierarchy
- assessment items aligned to it
- task, evidence, and learner models
- confirmation of the reliability and validity of items

Diagnostic assessment:

- placed before instruction
- items cover the range of skills, abilities, and difficulty levels
- decision-making accuracy of diagnostic confirmed

Formative assessment:

- placed within and between instruction
- design content for learning, not simply measuring
- incorporate a range of complexity depending on objectives and location in learning
- use primarily to pinpoint misunderstandings and provide feedback
- provide performance information (i.e., scores) to the instructor and/or to guide subsequent instruction

Summative assessment:

- placed after instruction
- make sure all learning objectives are assessed with at least two assessment items
- incorporate a range of complexity based on objectives
- use primarily to measure learners' proficiency level
- provide performance information (i.e., scores) to the instructor so they can evaluate and regulate their subsequent learning, and potentially improve instructional content

Domain model

The starting point for ECD is the domain model — a clear lay of the land of the learning domain that helps plan effective instruction and gives an accurate idea of a learner's knowledge level.

What it feels like for learners

I can tell that the specific thing I'm about to learn is a part of the wider field I want to get better at.

A domain is an area of performance, such as calculus or management. Domains consist of many subsets of knowledge, skills, and abilities.

The domain model articulates the knowledge, skills, and abilities that define success. This is important because it defines what we can assess in learners. A domain model is built of objectives or external frameworks, or both. They consist of dependencies and are hierarchical.

When beginning evidence-centered design, be clear on what domain you are trying to assess and what general claims you want to make about the learner and their learning.

What it means for designing learning experiences

- Identify either a subset of learning objectives or external frameworks or topics that are important within the domain
- Use input from appropriate subject matter experts. Ask:
  - What is important about this domain?
  - What work, tasks and situations are central in this domain?
  - What knowledge representations are central to this domain?
  - What objectives, behaviors, constructs, and attributes do we need to measure for evidence?
- Create a domain model, which can simply be an organized list of knowledge and skills or objectives
  - Include enough detail to diagnose and pinpoint learner challenges.
  - It can also represent misconceptions or affective states
- Organize this list of learning objectives or external frameworks hierarchically and articulate dependencies

See this Learning Design Principle:
Objective Design

Learner model

Showing and tracking the development of understanding in a domain, or learner proficiency, over time is needed to target instruction design and personalized learning. This is the learner model.

What it feels like for learners

> I can track where I'm proficient and what I still need to learn.

The learner model is a mathematical model of how learner proficiency changes over time. It maps or is a subset of the domain model. It is important because it answers the question: 'What exactly are we measuring?'

The learner model could look like:

- a subset of learning objectives and associated enabling objectives
- a subset of topics from an external framework
- a specific set of skills pertinent to a specific job-related task

The learner model is populated with specific learner data, and can then be used to make valid claims about a learner's proficiency in these objectives or skills

What it means for designing learning experiences

- Determine which parts of the domain model you want to track, assess and represent to learners
- Explore different ways of scoring learners' proficiency (e.g., percent correct or a more sophisticated method)
- Explore how proficiencies will be presented to learners in a dashboard
- Create the assessments based on the learner model
  - The task model and evidence model will fuel the learner model
  - The more detailed we want to get, the more instances of assessment we need — a minimum of three assessment points for each objective/skill
  - How can we structure the variables based on weights, difficulty, and dependencies?

Task model

Learning tasks or activities reveal evidence about what a learner knows and can do. Tasks should be designed to capture evidence of success, or learner behavior that reflects the desired objective or outcome. This is the task model.

What it feels like for learners

> By doing this task, I can find out if my learning is on track.

The task model specifies the tasks or activities that will prompt behaviors associated with the learner model, which allow us to make inferences about learners. It is important because it answers the question: How do we measure learning?

The task model can vary in its complexity, from simple selected response items to authentic simulations, with different trade-offs. The chosen task(s) must appropriately prompt and reveal evidence of learning.

Tasks can be simple as multiple choice or complex as navigating through a virtual world. Task models can look like:

- select-response questions
- writing assignments
- driving exams (real on-the-road or simulated)

What it means for designing learning experiences

- Design a set of tasks or activities to reveal evidence of what a learner knows or is able to do (i.e., that will prompt behaviors associated with the learner model)
  - The behaviors you specify in a task (i.e., the evidence) must be reliably measurable and clearly linked to one or more learner model variables
  - Tasks must reveal evidence that is granular enough to inform inferences about students
- If using selected response items, make sure there is more than one item per learning objective
- Select tasks that align with the purpose of the learning experience (e.g., a driving task to assess driving ability, recall items to assess knowledge of facts)
- Make sure evidence can be collected with the selected task models with current capabilities available
- You may need several task models to adequately meet the needs of the learning experience

Evidence model

All the ways a learner could respond to a task count as evidence and provide insights into their proficiency and understanding. This is the evidence model.

What it feels like for learners

The way I respond to this task provides insights into my understanding and abilities.

The evidence model lists all the behaviors prompted in a task and assigns weights to the learner model variables. It is important because it answers the questions:

- How do we measure it?
- What behaviors should reveal different levels of the objectives or competencies we're interested in?

Evidence models provide detailed instructions on how we should update our information about the learner model variables given learners' work on tasks. They contain two parts:

- evidence rules, describing how observable variables summarize a learner's performance in a particular task from the work they produced for that task
- measurement model, providing information about the connection between learner model variables and observable variables

What it means for designing learning experiences

When creating an evidence model, consider:

- What specific behaviors can we observe in the activities that inform the learner model?
- Can we align these behaviors with explicit skills or objectives in the learner model?
- What capabilities are necessary to capture observations, behaviors, performance?
- Have all behaviors relevant for each learner model claim been identified?

Example: driving exam

Learner model variable	Behavior	Score	Proficiency
Applies the stopping rules of driving	Completely stops at stop sign	1	50%
	Slows down when seeing yellow light	0	
Can accurately steer a car	Slows down before the corner	0	50%
	Passes cars with 2 feet distance	1	

The behavior column contains the evidence model claims: i.e., behaviors directly linked to learner model variables. Evidence scores are assigned to each claim, and the learner model is updated with proficiency scores.

Valid, reliable, and fair assessment

Assessments should consistently and accurately measure what they claim to. A reliable, valid, and fair assessment means we can make sound claims about learning.

What it feels like for learners

I can trust the assessment to measure what I know and what I need to learn.

Validity

- the degree to which evidence and theory support interpretations of test scores entailed by proposed uses of tests
- e.g., if the questions in a written driving test use lots of compound sentences, the test may end up testing reading comprehension rather than driving knowledge; this is not a valid assessment of driving knowledge

Reliability

- the consistency of such measurements when the testing procedure is repeated on a population of individuals or groups
- e.g., if a driving test takes place in a town where there's traffic during rush hour, tests that take place at midday and at rush hour will be quite different; this is not a reliable task for measuring driving performance

Fairness

- the degree to which an assessment produces similar results for different people, free of bias
- e.g., if a written driving test includes contexts that certain learners might find sensitive or not have as much knowledge of, this might not be a fair assessment of driving knowledge

What it means for designing learning experiences

Validity & reliability

- Ensure psychometric validity and reliability, use pre-existing instruments if available
- If claims and/or decisions will be made about learners based on the assessment data, consider conducting validity and reliability research, especially if high-stakes

Fairness

- Use assessment that supports learning and reflection, including formative assessment with feedback
- Design assessments that are open and linked to clear criteria (rather than relying on competition with others)
- Include a range of assessment strategies so that all learners have a chance to perform well

Making it inclusive

We aim to design a great learning experience *for everyone*. These considerations will help you get closer to designing a truly inclusive experience.

Accessibility

Accommodations provided to allow learners with specific needs to participate in a learning experience

- Make sure assessments of all types are accessible and have been tested using a range of assistive technologies. If an assessment is inaccessible, learners in need of accessible materials and technologies will be unable to demonstrate what they know.
- Encourage assessment creators to include strategies in formative assessments for putting insights about one's learning into manageable and actionable steps.

Availability and access

Barriers — often external and systemic — that affect whether the tools and resources needed for learning are available to learners

- Make sure the technology required is responsive (works on any device) and accommodates low bandwidth.
- Make sure learners are familiar with navigating the assessment in the environment it is being delivered in. Make sure there is time for learners to walk through each available technology before the assessment.

Identity

The combination of dimensions that define who learners are to themselves and others

- Mitigate identity threat for members of historically marginalized communities by providing opportunities for self-affirmation; to see mistakes as learning opportunities; and to reframe anxiety as not predictive of failure.

Culture

Shared patterns of behaviors and interactions, cognitive constructs, and affective understanding that learners acquire through socialization

- Encourage assessment developers to integrate options within the diagnostic to not privilege specific language and literacy skills if these are not what is being measured (e.g., if an assessment is not testing writing, allow learners to verbally respond to open-ended questions).
- Supply necessary background knowledge and micro-opportunities for skill-building, because prior education can vary greatly based on what is prioritized in an educational system within a culture.

In partnership with CAST

CAST is a nonprofit education research and development organization that created the Universal Design for Learning framework and UDL Guidelines, now used the world over to make learning more inclusive.

Authors

Katherine McEldoon, Ph.D.

Katherine McEldoon is a research-to-practice connector. After earning her Ph.D. in cognitive and learning sciences at Vanderbilt University and a post at Arizona State University's Learning Sciences Institute, she has worked in academia, government, and industry to ensure the best scientific insights support student learning, no matter the context. Katherine has most recently worked as Lead Learning Scientist on Pearson's Efficacy & Learning team, bringing evidence-based insights to Pearson's world of learners.

Matthew Ventura, Ph.D.

Matthew has 17 years' experience in the field of education technology research. Over his career he has managed multi-year research projects with large-cap technology companies and higher education institutions. Matthew's expertise spans research and assessment design, learning science, and artificial intelligence (AI) design for learning and assessment. He has over 70 publications in educational technology, next generation assessment, and AI.

Reviewers

Sara Finnigan, M.A.
Principal Learning Designer

Amy Wood, M.A.
Director, Learning Research and Design

Making It Inclusive

Autumn Westphal, M.A.
Learning Designer

Tanya Churaman, M.S.
Researcher

Selected references

- Ausubel, D.P. (1968). *Educational psychology: A cognitive view*. New York: Hold, Rinehart and Winston.
- Behrens, J.T., Mislevy, R.J., DiCerbo, K.E., & Levy, R. (2010). *An evidence centered design for learning and assessment in the digital world.* (CRESST Report 778). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Bejar, I. I. (2010). Application of Evidence-Centered Assessment Design to the Advanced Placement.
- Day, J.D., & Cordon, L.A. (1993). Static and dynamic measures of ability: An experimental comparison. Journal of Education Psychology, 85, 75-82.
- Hendrickson, A., Huff, K., & Luecht, R. (2010). Claims, Evidence, and Achievement-Level Descriptors as a Foundation for Item Design and Test Specifications. *Applied Measurement In Education*, 23(4), 358-377.
- Hoffman, A.M. & Wine, M. (2014). *Rigorous Test Development Introduction: The Core Principles*. Paper Presented at the Annual Meeting of the Northeast Educational Research Association. Trumbell, CT, 2014.
- Kibble, J. (2017). *Best practices in summative assessment.* Advances in Physiology Education, 41: 110 119. doi:10.1152/advan.00116.2016.
- Messick, S. (1987). *Validity. Educational Measurement* (3rd ed.) New York: Macmillan. Educational Testing Services.
- Mislevy, R.J., Almond, R.G. & Lukas, J.F. (2004). *A Brief Introduction to Evidence-Centered Design.* Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Sawyer, R. K. (2006). *The Cambridge handbook of the learning sciences.* Cambridge: Cambridge University Press.
- Shute, V., Kim, Y.J., Razzouk, R. *ECD*. myweb.fsu.edu/vshute/ECD.pdf
- Williamson, D. M., Bauer, M., Steinberg, L. S., Mislevy, R. J., Behrens, J. T., & DeMark, S. F. (2004). Design Rationale for a Complex Performance Assessment. *International Journal Of Testing*, 4(4), 303-332.
- Wiliam, D. (2018). *Embedded Formative Assessment* (2nd ed.). Bloomington, IN: Solution Tree Press.