Feedback Aide
AI, Responsibility, Security, and Privacy

Learnosity

The learning and assessment industry has entered an exciting new era of innovation. AI promises great change, but this provokes natural and logical questions: What changes? How will it impact us? What are the risks? How do we responsibly achieve change?

The questions we receive illustrate the curious, considered minds of an active community of product builders, authors, publishers, and assessment users.

We attempt to answer your most pressing queries here.

These questions are specific to the Learnosity AI product, Feedback Aide, but the responses here speak to our core approach to safeguarding user privacy, customer security, and responsible development across all our AI products. You can find similar information on our other AI product, Author Aide.

Since Feedback Aide is subject to continual improvement, this document is accurate as of the time of publication. If there's anything we don't cover in the following pages, please feel free to get in touch.

December 2024

Learnosity's mission and AI vision

Learnosity believes that education and learning opportunities should be equitable, inclusive, and accessible to all.

Our mission is to advance education and learning worldwide with best-in-class technology. As part of our mission, we're committed to the responsible use of AI in education and learning.

This document attempts to provide answers to common questions about Feedback Aide concerning security, privacy, and AI ethics.

Our AI manifesto

We believe that AI:

Reframes what's possible in assessment
AI gives us a once-in-a-lifetime opportunity to transform assessment so that educators, learners, and customers can focus on the things that matters most.

Should be used as a supplement, not a substitute
We view AI as a powerful tool for supporting and elevating human agency.

Should empower all
AI should be equitable, available to organizations of all sizes, and help meet individuals' diverse needs.

Should be used responsibly
AI systems must be secure, efficient, and reliable in order to protect user privacy and academic integrity.

Requires ongoing commitment
AI is an emerging field and we anticipate rapid and continuous change. We're committed to adapting to the uncertain by exploring the opportunities and risks that arise as the landscape changes.

Feedback Aide—an overview

Feedback Aide is an AI-enriched tool that supports human grading of essays and other unstructured question responses. Acting as a second reader, Feedback Aide lightens the workload for graders by taking a first pass at essay scoring before returning the reins for manual review and validation.

Feedback Aide is an embeddable API with an optional user interface that can be used to assess a learner response against a rubric. Learnosity's AI engine generates an initial score for a learner response, given the question stem and a grading rubric, as well as actionable, written feedback. Learnosity's FeedbackAide API then helps graders easily edit the generated scores and feedback.

A typical rubric and assessment criteria is provided below.

Explain the problems, both personal and societal, resulting from obesity

Obesity is a growing issue that many people, including children and adults, are facing in today's society. Obesity can lead to numerous personal and societal problems that affect individuals and communities as a whole. In this essay, I will discuss the different problems that result from obesity.

From a personal perspective, obesity can lead to several health issues such as diabetes, heart disease, and joint problems. For instance, excessive body weight can put a strain on the heart, making it work harder to pump blood throughout the body. This can then lead to heart-related problems, including heart attacks or strokes. Furthermore, obesity can cause type 2 diabetes, which is a condition where the body cannot properly regulate blood sugar levels. Additionally, obesity can result in joint problems because the extra weight puts

Criteria	① Unsatisfactory	② Needs Improvement	③ Satisfactory	④ Good	⑤ Excellent
Understanding of the topic	Fails to demonstrate an understanding of the topic	Demonstrates a limited understanding of the topic, with unclear or insufficient analysis	Demonstrates a basic understanding of the topic, with some basic analysis	Demonstrates a good understanding of the topic, with mostly clear analysis	Demonstrates a deep understanding of the topic, with clear and insightful analysis
Identification of Personal and Societal Problems	Fails to identify or analyse personal and societal problems resulting from obesity	Attempts to identify and analyse personal and societal problems resulting from obesity, but is unclear or incomplete	Identifies and analyses some personal and societal problems resulting from obesity, but lacks detail or clarity	Clearly identifies and analyses both personal and societal problems resulting from obesity, but may lack some detail or depth	Clearly and thoroughly identifies and analyses both personal and societal problems resulting from obesity

Feedback Aide is designed to assist and support human graders to evaluate and provide quality feedback. It is not a standalone grading tool.

Some key points concerning our approach to its development:

- Feedback Aide data is not used to train any public AI models.

- Feedback Aide is covered by Learnosity's ISO 27001 certification.

- In testing, Feedback Aide scores correlate highly with human grading in a wide range of subject areas.

- Feedback Aide is designed as an assistant to human grading and always should have a "human in the loop" to review the proposed score and feedback.

- Feedback Aide is designed to be scalable and robust, making it well suited for deployment in large scale scoring workflows.

- Learnosity acts as a data processor in its supply of Feedback Aide: the customer is always in control.

Feedback Aide—in depth

What functionality does Feedback Aide offer?

Feedback Aide is aimed at essays between 100 and 10,000 characters long, using AI to propose a score and feedback. In the event that the AI is not available or if the essays are too short or too long for AI scoring, essays can be scored entirely manually. There are different AI model approaches available, including a standard model and an advanced model, providing higher quality feedback. Learnosity will also shortly introduce moderation and flagging of inappropriate content and "cries for help".

Feedback Aide is under continual improvement, please see our documentation for full details.

How is Feedback Aide connected to other Learnosity / Questionmark software?

Feedback Aide is a standalone API. You can use Feedback Aide alongside the other Learnosity APIs or without them.

In the future, we expect to include Feedback Aide within the Questionmark platform to assist grading essays.

Where is developer documentation for Feedback Aide?

Follow the link to read our developer documentation on Feedback Aide.
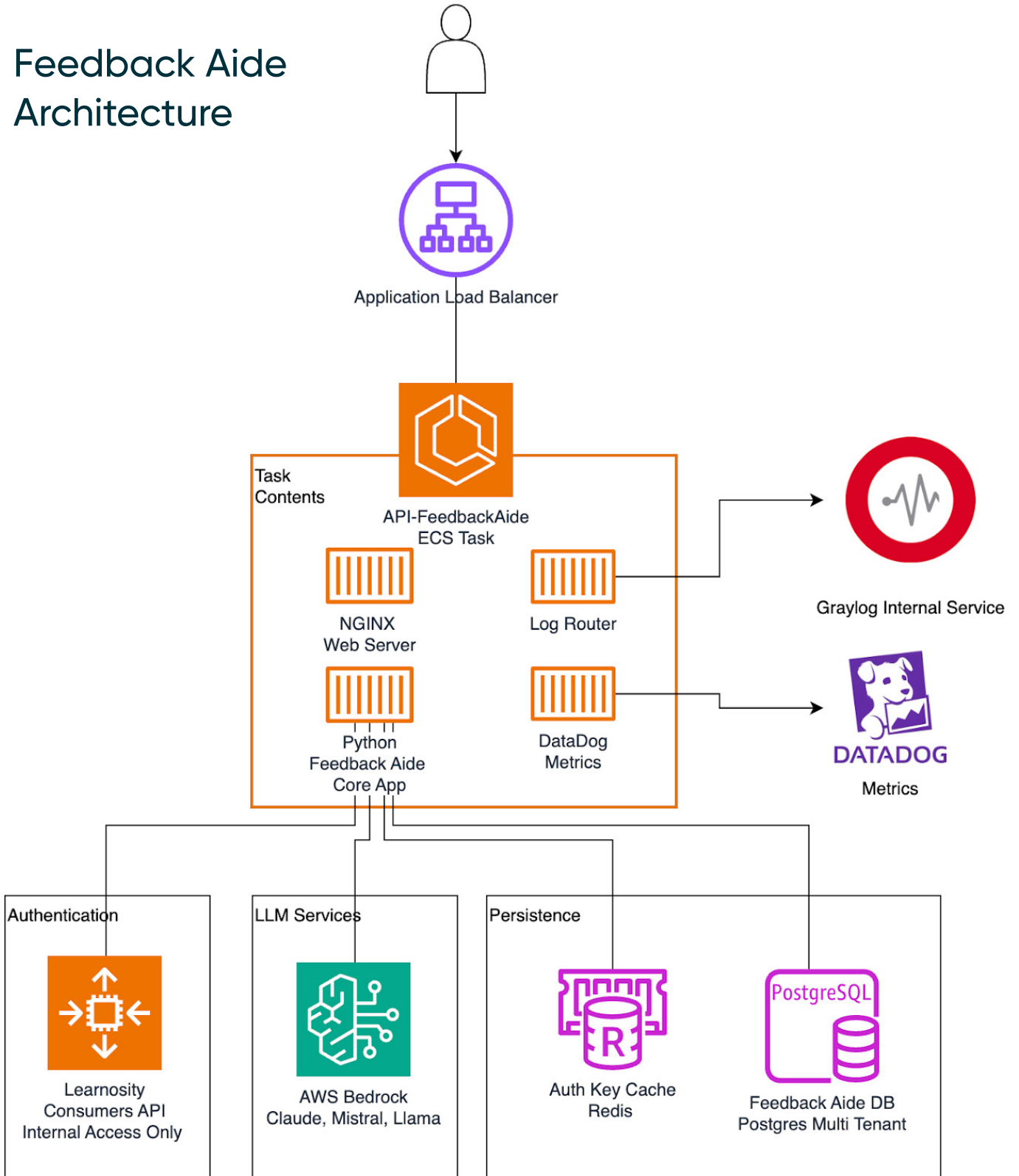
What is the architecture of Feedback Aide?

Feedback Aide is designed based on the 15+ years of Learnosity experience in building practical and scalable APIs for the education and assessment community. It is hosted on AWS and is designed to be a component that can easily be embedded in any system that provides rich essay scoring feedback.

What regions host Feedback Aide?

Feedback Aide is hosted on AWS, currently in the US in the us-east-1 region. Subject to customer demand, there is likely to be a European deployment in the future.

Feedback Aide Architecture

Application Load Balancer

Task Contents

API-FeedbackAide ECS Task

NGINX Web Server

Log Router

Graylog Internal Service

Python Feedback Aide Core App

DataDog Metrics

DATADOG

Metrics

Authentication

Learnosity Consumers API Internal Access Only

LLM Services

AWS Bedrock Claude, Mistral, Llama

Persistence

Auth Key Cache Redis

PostgreSQL

Feedback Aide DB Postgres Multi Tenant

Security

What certifications apply to Feedback Aide?

Learnosity's ISO 27001 certification applies to all our products, including Feedback Aide, demonstrating that we meet the highest standards for information security, including confidentiality, integrity, and availability.

You can view our ISO 27001 certificate here.

Learnosity is a member of the EDSAFE AI Industry Council, supports the SAFE Benchmarks Framework, and is a signatory of the Student Privacy Pledge. Our US companies (Learnosity Inc and Questionmark Corporation) are also self-certified against the EU-US Data Privacy Framework, its UK extension, and the Swiss-U.S. Data Privacy Framework.

Our hosting and AI providers are also 27001 certified and have several other certifications including SOC 2, ISO 9001 and ISO 27701.

What security measures have been taken while developing Feedback Aide?

Learnosity is committed to security, with a full-time security team that reports directly to our Chief Strategy Officer, a Learnosity co-founder.

We are externally audited as part of our ISO 27001 certification. We engage third party experts to conduct regular penetration tests on our products, permit customer penetration tests, and have other security audits.

Learnosity has a Product Development Security Policy as part of its ISO 27001-certified Information Security Management System. This policy is owned and managed by Learnosity's Chief Technology Officer and is reviewed and updated at least annually, the last update at the time of writing was April 2024. The policy covers areas such as access to source code, security training, use of third-party code, separation of environments, role-based training, and secure architecture principles.

Feedback Aide and all other Learnosity software are developed in alignment with our Product Development Security Policy. Learnosity follows a formal Security Development Lifecycle (SDLC) that includes information security considerations as well as privacy by design principles. This ensures that privacy and security are embedded throughout the development process. Our secure development process aligns with best practices, including OWASP.

How is access to Feedback Aide secured?

Feedback Aide uses an oAuth security pattern to handle authentication.

See the [developer documentation](#) for more information.

All communication with Feedback Aide is encrypted and all communication from Feedback Aide to the AI service is also encrypted.

For secure use of Feedback Aide, it's important also that the customer secures use of it appropriately, e.g. in how graders can use and how its software is protected from inappropriate access.

Is Feedback Aide vulnerable to injection attacks?

An injection attack is where special characters or code or AI prompts are included within other text seeking to influence or disrupt a system, for example to inject malware or to influence the output of the system.

Learnosity's designs minimize the possibility of this and we have a rigorous and expanding library of test cases that cover injection attacks.

How robust will Feedback Aide be under heavy load?

Learnosity has long experience developing scalable APIs, with over 19 billion questions delivered per year using its APIs. Load testing has been part of our work practices, and Feedback Aide is designed to be highly scalable and handle large loads. If you have particular queries, please ask us. We have also designed the software to gracefully degrade—if the AI model is not available for a request, graders can grade the essay manually.

Can I see further information on Learnosity's security?

Some information is at [General information on Learnosity API security](#). If you need more, please reach out to your business contact.

Privacy

What personal data is used within Feedback Aide?

Feedback Aide has very limited access to personal data:

- Learnosity does not know the identity of learners whose essays or other responses are scored by Feedback Aide.

- We also do not usually know the identity of graders.

- Feedback Aide processes the following learner data:
  - Essays or other answers from learners
  - Scores by rubric and for the essay as a whole
  - Feedback text given to the learner
  - Some technical data including IP address

- It also processes the changes to scores and feedback made by graders and date/time of scoring and other grading metadata.

In some jurisdictions, the data processed by Feedback Aide will not count as personal data as people are not identified. In other jurisdictions, including in the EU and United Kingdom, even though the data is pseudonymous, it is considered personal data.

Customers using Feedback Aide may also share with Learnosity personal information of administrators when raising technical support tickets (e.g. identity and contact data). Such information is stored confidentially and in compliance with privacy laws.

What role does Learnosity have under privacy laws?

Learnosity follows the instructions of its customers in how it processes their personal data, and is considered a processor, sub-processor, or service provider under data privacy laws.

Feedback Aide processes very little personal data, but any such personal data is treated securely. Contracts with customers and suppliers include data processing agreements.

Is there a Feedback Aide data processing agreement?

Yes—see https://learnosity.com/feedbackaidedpa.

Responsible AI

What generative AI model is used by Feedback Aide?

Feedback Aide is designed to be model agnostic and to be able to adopt models over time that best fit its intended purpose.

At the present time, the models used are Claude, Mistral and Llama hosted within AWS Bedrock, but other models are likely to be used in future. In addition, a mix of models will be used in some cases—either two models called for the same essay or else a model selected depending on the topic of the essay.

All models will either be hosted within AWS Bedrock or within Microsoft Azure Model as a Service. These frameworks protect the confidentiality of customer data and do not share customer data with the model provider.

Learnosity has contracts in place with AWS and with Microsoft Azure that ensure customer data is confidential and which forbid training on customer data. Data provided by customers to Feedback Aide is not shared with the model providers.

Do Learnosity AI providers use customer data to train their models?

No. Our AI service providers do not use any of the information provided to Feedback Aide or generated by Feedback Aide to train any model. The AI large language model is treated as "stateless" in this respect—i.e. not modified by Feedback Aide input.
Does Learnosity itself use customer data to train models?
There are two ways in which Feedback Aide can be purchased—Enterprise or Foundation.

Foundation is an entry-level license which allows Learnosity to use essay prompts, feedback, and scores anonymously to improve our models for all customers, including potentially training private (but not public) LLMs.

Enterprise does not permit use of data for training in any way.

Is there monitoring for abuse?

Our AI service providers do conduct some abuse monitoring. See AWS Bedrock abuse monitoring and Microsoft Azure abuse monitoring.

Learnosity is also adding moderation and safety alerts into the system, so that if a learner uses an essay to mention self harm, a cry for help, or sexual content, this can be flagged. Please reach out to your business contact if you want more information on this.

Who owns the copyright in material generated by Feedback Aide?

No copyright or other IPR is transferred to Learnosity by using Feedback Aide. Learnosity

does not claim any rights in questions or essay answers, other than those limited use rights necessary to provide Feedback Aide to the customer than owns or licenses those questions or essay answers.

Feedback Aide generates textual feedback using AI. Neither Learnosity nor our service providers make any claim to the copyright of the feedback text generated by Feedback Aide. Learnosity considers that customers own the output from Feedback Aide, as much as possible by the applicable legal jurisdiction[1].

Is any Feedback Aide data from one customer shared with another customer?

No. Neither Learnosity nor our AI service providers share any input to AI or output from AI with other customers.

If Feedback Aide is used multiple times with the same input, will it produce the same output?

No, not usually. Generative AI is typically non-deterministic, and repeating a call or regenerating will usually result in different outputs.

Can Feedback Aide performance drift or change over time?

In general, it is expected that Feedback Aide's performance will improve over time, as Learnosity improves the software, and as generative AI models themselves improve. However, generative AI model performance can change or drift, so it is important to monitor performance and use of Feedback Aide over time. We have a large bank of high-quality test data that we use to monitor performance and validate that the quality and accuracy is improving.

Does Feedback Aide use a chat interface to AI?

No. Feedback Aide calls the AI large language model via a secure API and controls its calls with carefully designed prompts that are intended to make it more likely that the scoring and feedback are useful.

What is the advantage of using Feedback Aide over having graders interact directly with a model like ChatGPT?

Feedback Aide allows a scalable workflow for using generative AI and is likely to be much more accurate and consistent due to the extensive testing and validation we have done. It is also more efficient and practical than direct use of generative AI models as it can be integrated into the grading flows.

---

[1] Note that in some jurisdictions it may not be possible to claim copyright on AI-created material.

Are the scores and feedback generated by Feedback Aide accurate and reliable?

In our testing with customer/partner datasets and with public datasets, the scores generated by Feedback Aide have a high correlation with those generated by human scorers, and the text feedback is generally appropriate and useful.

However, there is variation of performance depending in part on the subject area of the essay and in part on specific learner answers.

Feedback Aide is designed to give scores and feedback that are reviewed and improved by a human grader. It is designed to save time and resources in grading and to provide better text feedback than human graders typically have time for. However, occasionally there may be some inaccuracies for humans to correct.

Is it safe to share Feedback Aide scores and feedback with learners without human review?

Having a human in the loop is important, and we do not recommend sharing Feedback Aide scores and feedback with learners without human review at this time.

Does Feedback Aide explain the reasoning behind its scoring?

In general, generative AI works via statistical principles and is not capable of explaining the rationale behind what it produces.

However, because Feedback Aide works by giving scores at the rubrics level and includes feedback text for each rubric, the individual rubric scores (which add up to the total score) do give some explainability for the total score provided.

Is it necessary to inform learners and test-takers that AI is being used to assist scoring?

This will be a decision for each user of Feedback Aide to make and will be informed by organizational policy and applicable laws and regulation. Many regulations encourage transparency in the use of AI. For many organizations it will be appropriate to include in learner/test-taker communication that AI is used to assist human graders.

Can the scores and feedback from Feedback Aide be biased?

AI models are trained on a large corpus of human writing, which may itself contain bias. Learnosity prompts the AI to reduce some kinds of bias, but it is important that scores and feedback get reviewed by a human to help reduce the risk of bias.

However, human graders often have wide variances between them. And although most organizations have robust controls to protect against this, there is a risk that human graders themselves show bias. There is likely, therefore, a benefit to having an AI proposed score and

feedback that a human can review. It is a plausible hypothesis that scores and feedback generated by AI plus human review will be more inclusive and diverse than scores and feedback generated by humans only.

This is an area that we encourage customers to monitor and review over time, as will we be doing at Learnosity.

How inclusive and diverse is Feedback Aide?

Learnosity seeks to ensure via effective prompting that Feedback Aide scores and feedback take into account diverse learners and test takers. We do this by identifying, measuring, and iterating to build in safeguards, and reduce biases for particular issues.

What are the key benefits and risks of using Feedback Aide?

Benefits will vary by organization but are likely to include:

- Significant time-saving (and therefore money) in grading essays.

- Faster throughput so learners get scores and grades faster.

- Better quality and more thorough feedback to learners.

- Higher quality scores through the combination of AI and human grading.

- More consistency in scoring and feedback.

Risks will also vary by organization but potentially include:

- Risk of insufficient human oversight; human graders might consider that the AI is too capable and not thoroughly review scores and so miss outliers.

- Risk of poor input giving poor output. Poor rubrics will lead to poor feedback.

- Risk of bias in scoring.

- Failure to identify cries for help or other concerns in learner writing.

- Regulatory risk—need to ensure that use of AI is consistent with laws and regulations.

What do you recommend that customers consider when seeking to use Feedback Aide within a framework of responsible and ethical AI?

Some of the factors to consider will include:

- **Change management and training**
  The move to AI assisted grading is likely to involve some changes on how human graders are managed and trained.

- **Documenting use of AI**
  It's good practice to document all use of AI in assessment within your organization, and Feedback Aide should be included in this.

- **Efficacy**
  Does Feedback Aide improve the efficiency of question and test creation? It's worth considering if you can obtain qualitative or quantitative measures to help evidence this.

- **Human oversight**
  It is important that there is genuine human oversight and review of the scores and feedback to deal with errors or bias by the AI.

- **Legal compliance**
  Many jurisdictions are introducing laws in relation to use of AI, and these need to be monitored. Also, many existing laws apply to AI as they do to non-AI technologies.

- **Review**
  AI technology is changing quickly - put in place a review cycle to review your use of Feedback Aide and other AI.

- **Supplier contracting and vetting**
  Ensure that you have appropriate contracts with suppliers, including Learnosity.

- **Transparency**
  Most organizations will want to share with learners their use of AI.

For more information on ethical and responsible AI within assessment, see "Creating Responsible and Ethical AI Policies for Assessment Organizations" published by the ATP in 2024, which had significant input from Learnosity.

About Learnosity

Learnosity is the global leader in AI-optimized assessment solutions.

Serving over 750 customers and more than 40 million learners, our mission is to advance education and learning worldwide with best-in-class technology.

More at **learnosity.com**

Learnosity