ICT Project Guidance

Education Sector   
Assessment System Aspects to Consider

Version:

0.5

Author:

Sky Sigal

## Description

This document outlines some dimensions to consider when discussing desired capabilities of an assessment system.

## Synopsis

Assessment systems options range from emulating traditional proctored experiences to permitting assessments be taken anywhere, on any device, at any time, many times without adding risk of generating unwanted outcomes.

## Contents

[Description 1](#_Toc148453854)

[Synopsis 1](#_Toc148453855)

[Contents 2](#_Toc148453856)

[Objective 4](#_Toc148453857)

[Risks 4](#_Toc148453858)

[Resolution 4](#_Toc148453859)

[Considerations 5](#_Toc148453860)

[Intended Assessees 5](#_Toc148453861)

[Assessment Areas 5](#_Toc148453862)

[Device 6](#_Toc148453863)

[Quality Assurance 6](#_Toc148453864)

[Assessment Location 7](#_Toc148453865)

[Assessment Event Scheduling 7](#_Toc148453866)

[Assessment Item Selection 8](#_Toc148453867)

[Assessment Item Steps 8](#_Toc148453868)

[Assessment Item Step Roles 9](#_Toc148453869)

[Assessment Response Types 9](#_Toc148453870)

[Assessment Duration 10](#_Toc148453871)

[Assessment Item Duration 10](#_Toc148453872)

[Assessment Scoring 11](#_Toc148453873)

[Assessment Scoring Event 11](#_Toc148453874)

[Assessment Item Complexity 11](#_Toc148453875)

[Inline Assistance 12](#_Toc148453876)

[Assessment Scheduling 12](#_Toc148453877)

[Experience 12](#_Toc148453878)

[Assessee Roles 13](#_Toc148453879)

[Integrations 13](#_Toc148453880)

[Subject Depth 13](#_Toc148453881)

[Subject Breadth 13](#_Toc148453882)

[Test Media 14](#_Toc148453883)

[Test Response Media 14](#_Toc148453884)

[Feedback Collection 14](#_Toc148453885)

[Feedback Source 14](#_Toc148453886)

[Item Development 15](#_Toc148453887)

[Standardisation & Calibration 15](#_Toc148453888)

[Reporting Audience 15](#_Toc148453889)

[Reporting Options 15](#_Toc148453890)

[Lifespan Purpose 16](#_Toc148453891)

[Platform Considerations 17](#_Toc148453892)

[Appendices 18](#_Toc148453893)

[Appendix A - Document Information 18](#_Toc148453894)

[Images 18](#_Toc148453895)

[Tables 18](#_Toc148453896)

[References 18](#_Toc148453897)

[Review Distribution 18](#_Toc148453898)

[Audience 18](#_Toc148453899)

[Structure 18](#_Toc148453900)

[Diagrams 18](#_Toc148453901)

[Terms 18](#_Toc148453902)

## Objective

The objective is to develop one or more assessment services for the sector that are valued for delivering insight into individual learners’ achievements and support needs that leads to improving learning outcomes.

## Risks

Working on legacy platforms with outdated requirements has the risk of limiting thinking of potential objectives that could be made possible with more current system design options.

## Resolution

This document provides a list of outcomes to consider when developing a list of desired outcomes from a service. The list of desired outcomes in turn leads to developing tasks to modify existing systems or the development of requirements for new systems.

# Considerations

The following categories are continuums to consider.

### Intended Assessees

Assessment services adjust their interface, complexity and content to address different markets: primary, intermediate, and secondary learners, or professional adults being assessed for PLD or enterprise onboarding objectives.

Options to consider include the following:

* National compulsory learning:
  + Primary learners
  + Intermediate learners
  + Secondary learners
* Professional Learning Development (PLD):
  + Provider Teachers,
  + Provider’s other Workforce (e.g.: Admin, etc.)
  + Sector workforce (e.g.: Agency members)
  + Ministry enterprise learners

### Assessment Areas

While the immediate is important, consider selecting assessing system with capabilities of delivering immediate needs -- while being incrementally improvable to delivery of current needs, and/or meet future needs.  
  
Consider the following example options as longer term objectives:

* Maths (primitive quiz capabilities)
* Maths, Comprehension (requiring prompt media)
* Maths, Comprehension, and Writing (requiring textual input and analysis)
* Maths, Comprehension, and Writing in multiple languages (categorisation of assessments by culture and language – implying UI changes)
* Maths, Comprehension, Writing, Natural and Social Sciences (wider selection of tests, more complex categorisation & search, etc.)
* Maths, Comprehension, Writing, Natural and Social Sciences, Individual Arts (richer input of images, voice samples for analysis, potentially even video analysis by AI)
* Maths, Comprehension, Writing, Natural and Social Sciences, Group Arts (potentially multi-role, combined with group activity video analysis).

### Relative priority of Paper Based Assessments

Digital services may not be available for a number of reasons:

* Travel may not be available from remote locations (e.g.: home schooling) to a web accessible location,
* Devices (e.g.: BYOD) may not be charged, and there may not be sufficient power to run all devices brought to a common assessment location (e.g.: school hall, etc.)
* Assessment locations (e.g.: remote homes and/or school gyms) not have sufficient the capability to provide sufficient wired or wireless network connectivity to devices.
* A number of users may refuse to use digital services, and/or do not have required digital Ids (e.g.: for religious or other reason).
* Unexpected peak demand may lead to accessibility and/or throughput issues, sufficient to require aborting taking the assessment.

Although all the above cases are marginal in quantity, the ability to print a paper-based assessments can be employed to provide Business Continuity.

But Paper is a static format, at odds with dynamically responsive assessments.

It is therefore important to determine early whether the system is to be designed as a paper-first based solution, so that all users, whether online or not, have the same questions, or the flexibility of dynamic adaptability is more important, with only a subset falling back to a pre-fixed test sequence.

Note:  
In essence the question is whether the desired outcome is assess against a test, or a domain.

### Device

While still uneven in penetration of schools due to ideological, religious and financial reasons, digital devices are here to stay and increase their use in every aspect of teaching.

Bring Your Own Device (BYOD) is an increasingly accepted reality: learners are bringing their mobile devices for personal reasons: it is logical to leverage these devices to decrease funding required for desktop devices.

Options to consider include the following:

* performed on any web-enabled device, whether BYOD or school provided, with no software installation prerequisites. Adaption of assessment information is required to address acceptance that users may use the internet to search for answers.
* Performed on a web-enabled device, whether BYOD or made available by the education provider, with a prerequisite for the installation of software to monitor their behaviour (e.g.: web traffic) to limit the use of the web.   
  Note that this impacts privacy concerns.
* Performed on a device, whether BYOD or made available by the education provider, on which software must be installed, that disables web access. This impacts the ability of the system to be dynamic in its choice of questions if it cannot call back to a central service.
* performed on devices provided by the assessor’s organisation, specific to the task, that preclude the assessee having previously saved material that would help.   
  Note this has large funding implications, impacting rollout by such a duration that the service design and capabilities may be outdated before fully implemented.
* Preformed on devices that are locked down to not allow access to other software (browsers, etc.) through which access to resource material could be obtained. Same implications.

### Assessment Area Integrations

Whereas western societies value assessments of single topics, other cultures value more the cohesivity of topics.

For example, while European cultures may be used to taking courses and assessents of physics, chemistry, and biology separately, other cultures may prefer assessing the three topics as simply aspects of a larger entity, nature itself.

Stimuli media may be developed to test either for one assessment area (e.g.: mathematics) or several areas, with stimuli derived questions on mathematics and other subjects, etc. all working together).

***TODO:*** *requires better discussion and demonstrations to be made more accessible to western culture teachers.*

### Quality Assurance

Proctoring of assembled assessees requires the attention, time and therefore cost of one or more persons for the duration of the test. The attention required removes opportunity to do other tasks. Proctoring can be automated, or its need be removed entirely, releasing time and effort for other activities.

Options to consider include the following:

* Manual: In person, on-site proctoring,
* Assisted: In person, technology assisted proctoring (e.g., Multiple video cameras, etc.)
* Automating Proctoring (by the use of software installed on the Device that control the laptop camera, web access, etc.), or AI managed cameras viewing assembled assessees, etc.   
  The requirement of an on-device camera – or camera peripheral – implies a necessity of funding a national rollout, that by the time it is completed to a sufficient level, the devices would out of date.
* Non-Proctored: based on adjusting content (e.g.: ensuring questions are randomly ordered, so no two questions #9 are the same, or randomly choosing questions from a bank, such that the questions are targeted at the same level, but different, etc.). This removes the need for the distribution of a specific device, or use of a specific location.

### Assessment Location

Traditionally, Assessees have been required to assemble in one location, for proctoring by staff. The location may require an upgrade to Wi-Fi networking capabilities to handle the larger than average number of connections required.

Options to consider include the following:

* Onsite, proctored,
* Onsite, non-proctored,
* Blended, permitting a mix of people on site, as well as remote, having been provisioned beforehand with any specific device.
* Remote, permitting users from anywhere, having been provisioned beforehand with any specific device.
* Remote, un-monitored, requiring no specific device, having adjusted content and selection automation.

### Assessment Event Scheduling

Traditionally, assessments are set at a specific time, with a specific deadline for completion. This used to require presence, and sometimes transport coordination (if done outside school hours). This negatively impacts some learners who can’t make that time (e.g., migraines, other distracting pain, family events including burials, etc.)   
Setting up a second event to test the above learners requires a duplication of effort.

Options to consider include the following:

\* Synchronous (all users take test at a specific time),

\* Blended: users can take tests at any time within a window (e.g.: over the following weekend is an acceptable delay, but no longer).

### Assessment Item Selection

Traditionally, items were pre-selected by a teacher making a best effort to select questions that best reflect an intended level, and that haven’t been asked before. The effort to select items, even if assisted with an item bank of questions to select from, is considerable. It also limits testing to an average band, providing little guidance for learners that can answer most questions, and simply failing learners that don’t fit in the band.

Automation can remove this effort.

Options to consider include the following:

* Manual development of assessment questions and answers,
* Manual choice from pre-developed sets of assessment questions and answers,
* Manual choice of pre-developed sets of assessment prompts, with the ability to edit a percentage of them.
* Automated by Randomness,
* Automated by Randomness without Repetition,
* Automated by Past entries without Repetition,
* Automated by Past entries, without Repetition, adjusting downwards or upwards to provide better insight into the true level of edge learners.

### Assessment Set

Are Items only accept defining the correct outcome? Or are Items defined by Errors? Or both?

### Assessment Item Valid Distractors

The Errors may be used to list specific issues – valid distractors and similar -- that come up as often misunderstood.

### Assessment Item Steps

Developing items in steps. For example, early learners may only require two steps (Question/Answer), whereas intermediate learners may be expected to solve problems in specific steps, whereas later learners are supposed to know multiple strategies, each using different steps, but demonstrate the ability to choose the least costly approach.

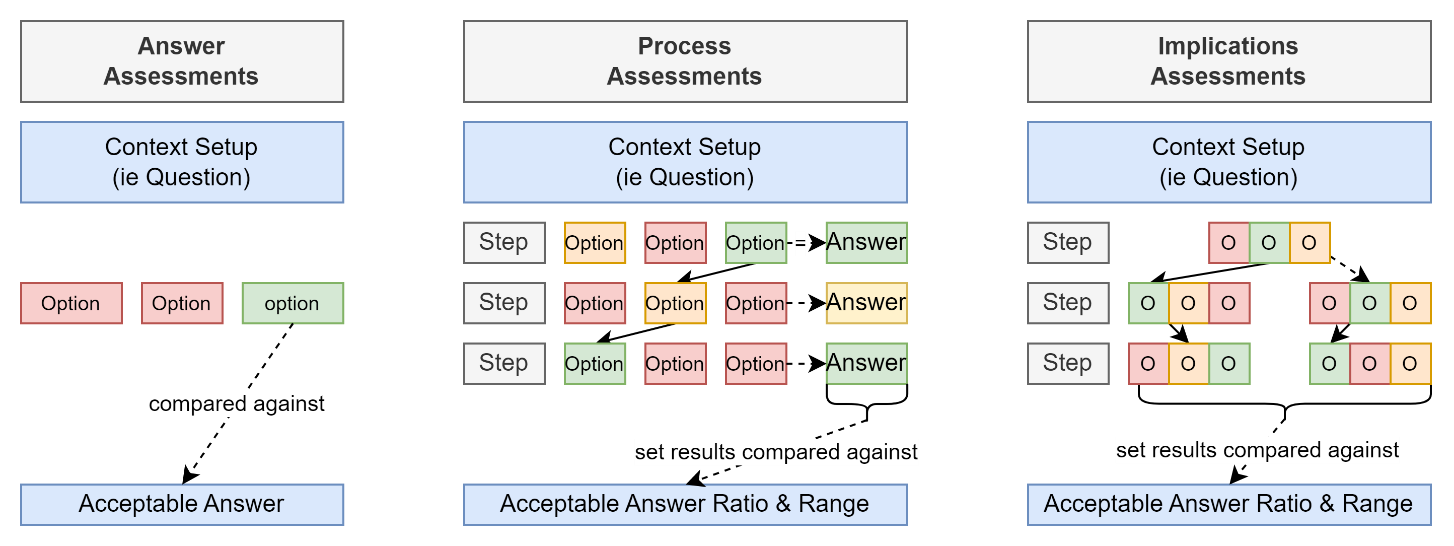


Figure : Sequential Assessment options

Options to consider include:

* Simple two step approach (Question/Answer)
* Multiple steps, at each step selecting one from a list of options.
* Multiple steps, permitting showing an order.
* Multiple steps, where the user enters more work, rather than selecting from an options table, and it in turn selects what to offer next. It’s interactive.
* Etc.

### Assessment Item Step Roles

Whereas in the western world assessments are primarily developed as a development of an individual’s ability, some cultures will always prefer weaving in an underlying understanding that knowledge for the sake of knowledge, separate from the ability to work with others to apply it, is a simulacrum only that does not provide as valuable an assessment.

Options to consider include:

* Simple, traditional, roles (Assessor, Assessee)
* Multiple Roles can work on all Steps,
* Roles per Step.

### Assessment Response Types

The assessment areas, along with assessees affects may impact what kind of assessment can be done.

Options to consider include the following:

* Selection:
  + Binary Options (Yes/No)
  + Single Options from List (Radio Buttons)
  + Multiple Options (one or more selected from the list)
  + Value (slider)
  + Range (bottom and top value along a slider)
* Categorisation:
  + Order (re-arranging items of a set)
  + Mapping (from left list to right list)
* Input:
  + Location (X/Y over an underlay image, such as a map)
  + Uploaded media
  + Single line text entry
  + Multiple line text entry
  + Document

### Assessment Duration

Options to consider include the following:

* Time-bound
* Time-bound with earnable extension
* Time-unbound, taking as much time as is required to demonstrate achieved ability -- by correctly answering *n* (e.g., 5) times different questions on the same concept -- or abandoning.

### Assessment Item Duration

While an Assessment can be expected to be completed within a set time (e.g. an hour), the questions within may be of varying complexity and take more or less time to answer each one.

Options to consider include the following:

* None, falling back to the overall assessment duration, if there is a limit.
* Questions are all given the same amount of time (e.g.: 1 minute) impacting some users more than others (e.g., visually impaired users, users with dyslexia or similar, etc.).
* Individual Assessment items each have expected complexity (e.g., A, B, C durations which can be mapped to configurable times: 30 seconds, 1 minute, 2 minutes, etc).

### Assessment Scoring

Assessing the results (e.g., scoring) used to be done manually, after the return of printed test sheets. This can be automated to reduce workforce effort and time requirements.

Options to consider include:

* Manual: where scoring is by the teacher or other role after a test has been completed.
* Blended: some assessments can be automatically scored, such as maths using questions and multichoice answers, while others may require manual scoring, such as written responses.
* Automated: where all tests are graded by automation.

### Assessment Scoring Event

Scoring used to be done manually, after all assessments had been completed. Automation can remove the effort required, as well as remove the need for coordination of assessees to start and complete by a deadline.

Options to consider include the following:

* Post-completion: scoring is performed only after a person has completed or abandoned an assessment.
* Inline: scoring is changed after each question is submitted.   
  The user can be released when it becomes evident they won’t pass a minimum threshold even if they answered all remaining questions correctly.  
  Note this approach constrains users from being able to return to questions to correct them.
* Inline: each question informs the system what to ask next, but scoring is deferred to the end.

### Assessment Item Complexity

Options to consider include the following:

* The user has no control over the complexity of the question asked.
* The user is in control of asking for easier questions on the same context, accepting that it impacts scoring to some extent.

### Inline Assistance

This aspect is directly associated to several other aspects of an assessment service, including Purpose. For example, a tool that is only intended for assessment only needs to provide test questions and potential answers, whereas a tool that is intended to also be tutor users, or get them over the line by any means, may rely on assistance.

Note: Onboarding PLD for example is more focused on helping users achieve baseline results and benefit from either being permitted to redoing questions or having access to learning material between *“redo”s* or even while a question is being pondered.

Options to consider include the following:

* No hints allowed (closed book)
* Hints that cost (accessing hints would cost some points)
* Open book, with access to learning resources at no cost.

Note:  
Even if initial assistive material is not delivered in a first tranche, if the service doesn’t preclude this use case, the assistive material can be developed later.

### Assessment Categorisation: by Curriculums

Assessments can be categorised by alignment to curriculum levels.

Many countries have multiple curriculums.

There is an education sector protocol for importing Curriculum information: CASE.

### Assessment Scheduling

Options to consider include the following:

* Taking Assessments before learning, to inform the development of teaching plans  
  (“Assessments for Learning”)
* Taking Assessments during learning, to inform corrections to teaching plans (“Assessments for Learning”, potentially “Assessments as Learning”)
* Taking Assessments after teaching, to test the comprehension of the concepts (“Assessments of Learning”).

### Experience

Options to consider include the following:

* Text only questions and answers
* Media, media description and text-based questions and answers,
* Interactive & immersive 3D experiences

### Assessee Roles

Options to consider include the following:

* Simple role set: (e.g., Assessor and Assessee role)
* Simple interactive role set: (e.g., inline Asker/Guide, or leading/responding)
* Medium role set (e.g., Assessor, Assessee, Proctor, judgment Reviewer)
* Multi-role set (e.g., multiple Assessees can collaborate to solve query).
  + e.g., in NZC, Chemistry
  + e.g., in MM, Caller & responder, team efforts, etc.

### Integrations

Options to consider include the following:

* None
* Minimal ministry integration (e.g., IdP, SMTP email)
* Medium ministry integration (e.g., IdP, SMTP, Te Rito, Data Warehouse)
* Full ministry and platform integration (e.g., same, plus integration to other assessment services, and combine holistic reporting, potentially via BI)
* Sector (e.g., same, plus provider SMSs)

### Subject Depth

options to consider include the following:

Year level:

* Primary
* Intermediate
* Secondary
* Tertiary
* Professional

### Subject Breadth

Options to consider include the following:

* Comprehension
* Maths
* Writing
* Science
* Arts
* Dance & Gym (requires roles & video capture?)

### Test Media

Options to consider:

* Text
* Image
* Sound
* Video
* 3D

### Test Response Media

Options to consider include the following:

* Multichoice
* Text input
* Sound capture
* Photo (for later review or AI capture)
* Video capture (for later review or AI capture)

### Feedback Collection

Options to consider include the following:

* None.
* Inline: responding to specific questions
* Summative: feedback given at end of tests

### Feedback Source

Options to consider include the following:

* From Teachers only
* From Teachers and Learners
* From Teachers, Learners and Caretakers

### Item Development

Options to consider include the following:

* None
* Manual
* Pre-automated by automation, with Human in the Loop for QA and approval.
* Responsively: Inline automation responding to human interactions

### Integration: CASE Standard

TODO: Integration with Curricula standards (via CASE standard).

### Integration: QTI Standard

TODO: Not especially impressed with QTI, but it is still the most referenced standard available. For now.

### Standardisation & Calibration

Options to consider include the following:

* None
* Manual standardisation
* Automated standardisation adjusting based on usage.
* Removed (based on getting to level above goal standard) <- check!

### Reporting Audience

Systems can report to a range of stakeholders.

Options to consider include one or more of the following:

* Learners
* Teachers
* School Admins
* Parents

### Reporting Options

Reporting can include multiple information aspects:

* Actual outcomes:
  + Actual outcome (summative)
  + Actual outcome (detailed, including sections of assessment)
* Trend outcomes
  + Plotting expected outcomes over time, based on previous results,
  + Plotting expected outcomes compared to others in same class, school, region, country.
* Relative outcomes:
  + to other Learners results: Same class, same school, regionally, nationally, internationally.
  + Relative to Curricula: curricula’s intended progression levels, other national curricula, international curricula,

### Lifespan Purpose

The cost of developing automation, along with the development of the information contained within the system is considerable.

This leads to a need for services to have a very long service lifespan, that can negatively impact end users if it cannot be incrementally evolved. For example, some current systems were designed before everything that learners take for granted now: touchscreens, mobile phones, mobile data, electric vehicles, self-driving vehicles, drones, uber or uber-eats.

Note:   
It is not conjecture, it is a predicable outcome that children born today will be schooled during an era where they will encounter daily self-driving cars, drone and robot-based delivery, robots of various abilities within the workforce visible to the public (as opposed to currently hidden in factories), omnipresent cameras in some public areas, AI assistance and presence in lobbies, schools. They should expect to use systems that are not significantly incongruous with these comparable technologies.

A second strategy to maintain an acceptable value/cost ratio, may be required: in addition to being evolvable it for its *current* capabilities to remain current, it may also require being able to incrementally expand its capabilities to address *future* purposes.

For example, at a minimum, could a system designed for assessments only be developed further to become a tutoring service, accessible anytime, anywhere, from any device, by anyone?

Options to consider include the following:

* Assessments only
* Assessments with included hints
* Assessments and Tutoring service, using a subset of the item bank (to leave items for assessment only purposes)
* Assessments and Tutoring service, using any item from the item bank, permitting dedicated learners to learn at their own pace.

### Platform Considerations

A unified platform, capable of hosting an array of different types of assessments (for example, one module offering Teacher developed Overall Teacher Judgements, and another module capable of developing Learner developed assessments results) has both advantages and disadvantages.

For example, a unified platform may lead to less integration requirements, improved discoverability, easier teacher and learner onboarding, less training, and some other positive outcomes.

But it may also inadvertently impose a technological limit to incrementally incrementing to meet evolving needs that become required, but the base platform can’t deliver.

Options to consider include the following:

* Point solution, doing a single type of assessment, not integrated to support services, therefore requiring its own capabilities for printing, etc.
* Point solution integrated to reporting platform (e.g., integrated by Te Rito),
* Point solution, integrated to schools SMSs directly (and then to Te Rito).
* Common Platform (e.g., one of a set of modules sharing a common base platform)

Appendices

Appendix A - Document Information

### Images

[Figure 1: Sequential Assessment options 8](#_Toc148453820)

### Tables

### References

**There are no sources in the current document.**

### Review Distribution

The document was distributed for review as below:

|  |  |
| --- | --- |
| Identity | Notes |
| Sky Sigal, Solution Architect |  |
| Sandy Britain, Enterprise Architect |  |
| Rodney Snell, Team Lead |  |

### Audience

The document is technical in nature, but parts are expected to be read and/or validated by a non-technical audience.

### Structure

Where possible, the document structure is guided by either ISO-\* standards or best practice.

### Diagrams

Diagrams are developed for a wide audience. Unless specifically for a technical audience, where the use of industry standard diagram types (ArchiMate, UML, C4), is appropriate, diagrams are developed as simple “box & line” monochrome diagrams.

### Terms

Refer to the project’s Glossary.

##### IT

: acronym for Information, using Technology to automate and facilitate its management.

##### ICT

: acronym for Information & Communication Technology, the domain of defining Information elements and using technology to automate their communication between entities. IT is a subset of ICT.