NEEDS ASSESSMENT

The State of the Art in innovative assessment approaches for VN FIL
# TABLE OF CONTENT

**Introduction** ........................................................................................................................................ 4

1. **Progress in implementation of VNFIL: A brief overview** ............................................. 5

2. **Validity and reliability of assessments for VNFIL** ................................................... 7

3. **Priorities for development** ................................................................................................. 13

4. **A Summary of Needs for Innovation in VNFIL** ................................................................. 21

**References** ....................................................................................................................................... 22
Introduction

Europe has set the goal to make lifelong learning a reality for more people through creation of flexible pathways that improve access to higher levels of learning and employment for individuals who otherwise would have been denied these opportunities. To achieve this broad goal, a European Council Recommendation of 2012 urges countries to have national arrangements for validation of non-formal and informal learning (VNFIL) by 2018. At the policy level, a number of countries have made significant progress in developing frameworks and strategies to support VNFIL. These arrangements have frequently included introduction of alternative assessment methods for VNFIL such as portfolios, simulations, interviews, and so on, which are often seen as more appropriate for disadvantaged learners.

In spite of this important progress in introducing alternative assessments, take-up has lagged behind. As noted in Cedefop’s 2014 Synthesis Report, standardised tests, which are considered as being the most valid and reliable (the current gold standard for assessment), are still the most accepted methodology for VNFIL. Moreover, as VNFIL approaches are already outside the mainstream, formal education system, there are concerns that alternative assessments may create “B class certificates” (Cedefop, 2008).

The focus of this Needs Assessment is on effective assessments that better support individuals in disadvantaged groups to participate in VNFIL. Alternative assessments are seen as particularly important for individuals who may avoid any type of assessment which brings back memories of school failure (Tsekoura and Giannakoupoulou, 2017), or for which the language is too difficult. The report is aimed at practitioners (as opposed to policy-makers), and thus complements the Cedefop Guidelines and Inventory. It explores some innovative ideas, methods and tools that aim to improve the validity and reliability of alternative assessments for VNFIL. It should also be noted that alternative assessments are of relevance for all VNFIL candidates. Well-designed performance assessments typically allow the individual to demonstrate complex problem solving processes in a realistic context, while more traditional standardised assessments feature recognition and recall but do not typically capture how that knowledge is applied to address problems in a realistic context (Jonsson and Svingby, 2007).

The report complements two companion reports developed for the InnoVal project, including a report exploring the perspective of VNFIL candidates, and an online stakeholder consultation (Uras, 2017: The stakeholder consultation yielded 83 responses from 19 different countries. The majority of respondents are based in the countries of the project partners. They are PT (14 responses), BE (10), GR (10), FR (8), UK (6), DE (4), IT (4), DK (30), NL (4), SL (3), ES (3), IE (3), NO (2), RO (2), RO (2), FI (2), CY (1) LT (1), MT (1), Russia (1), EU organisation (1). Of these 36.1% were from the HE sector, 20.5% from adult learning, 6% from the occupational training sector, 6% from a public authority or government, 6% from employment counselling and 6% from the non-formal/informal learning sector.). It also sets out areas that will be of particular interest for the InnoVal case studies, which will be developed following publication of this report.
In the next section (section 2) we briefly set out progress in the policy development and implementation of VNFIL. In section 3, we examine the pros and cons of different assessment methods, and the extent to which they are seen as valid, reliable and usable for VNFIL. We also explore how a mix of assessment standardised and alternative methods may be used to develop a more complete picture of a candidate’s competences.

In section 4, we focus on priorities for development of alternative assessments of VNFIL, as identified in our stakeholder consultation and the literature. These include the importance of improvements to assessment tools and frameworks, capacity building for assessors, and cost effectiveness and quality assurance of the overall system. We also note feedback from stakeholders on the need for greater involvement in development of tools and approaches for VNFIL.

Finally, in section 5, we conclude with a summary of the main issues identified in this Needs Assessment, and which will be important for the InnoVal case study research.

### 1. Progress in implementation of VNFIL: A brief overview

Across Europe, countries have made clear progress in developing national validation policies and frameworks. National strategies for validation are typically integrated within broader education strategies/policies (e.g. Finland). In at least six countries (the Czech Republic, France, Italy, Portugal, Spain, and Norway), the national strategy for validation is outlined in legislation (Cedefop 2014).

The number of countries lacking a national strategy for validation has decreased from seventeen in 2010 to nine in 2014. Some countries with a national strategy have not enacted them; and some not currently with a strategy are in the process of preparing one, for example Austria, Cyprus, Denmark and Portugal. There are gaps in some countries’ national strategies, which include: insufficient support for take-up of validation, and/or low visibility of the process, a low involvement of the voluntary sector, and weak links between validation activities in the public, private and voluntary sector. In addition, country strategies for different sectors of education and training may not be integrated, and the coexistence of different validation practices in the same country makes data collection difficult on an aggregate basis (Cedefop Inventory VNFIL 2014).

<table>
<thead>
<tr>
<th>Comprehensive strategy in place</th>
<th>Strategy in place but some elements missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>FI, FR, ES</td>
<td>CZ, DK, EE, IT, IS, LU, LV, NO, NL, PL, RO</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy is in development</th>
<th>No strategy in place</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT, BE-Flanders, CH, CY, DE, EL, LI, LT, MT, PT, SI, SK, TR</td>
<td>BE-Wallonia, BG, HR, HU, IE, SE, UK- E&amp;NI, UK-Scotland, UK-Wales</td>
</tr>
</tbody>
</table>

Source: Cedefop, 2014
INNOVAL NEEDS ASSESSMENT

Most countries have developed sector-wide strategies for some education sectors (HE, VET, adult education, etc.) but not all. In some cases, sector-wide validation strategies exist for VET and/or adult education, but not necessarily for higher education (or only apply locally). In Estonia developments are more advanced in HE than in other sectors, while certain countries are developing sector-wide validation strategies, for example Latvia (VET and HE), Hungary (adult learning, HE), Italy (HE, VET), Sweden (adult education) and Scotland (HE) (Cedefop European Inventory VNFIL 2014).

Four main clusters of countries with legal frameworks can be identified – those with:
- a single legal framework for validation;
- multiple validation frameworks covering different sectors
- no legal framework for validation;
- legal framework for other initiatives, also covers validation; (for example in Iceland, the Adult Education Act has provisions on individual entitlement to the validation of non-formal and informal learning towards credits/units at the upper secondary level)

Table 1.2: Legal frameworks for validation

<table>
<thead>
<tr>
<th>Single legal framework for validation</th>
<th>Multiple frameworks in place for different sectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR, MT, TR</td>
<td>AT, BE (Flanders and Wallonia), GB, CH, CZ, DK, FI, ES, EE, DE, IT, LT, LV, LU, NL, NO, PL, SE, SI</td>
</tr>
<tr>
<td>Legal frameworks for other initiatives also covers validation</td>
<td>No legal framework covering validation</td>
</tr>
<tr>
<td>IS (Adult education), IE, HU (HE, Adult education, PT (HE and non HE), RO, SK</td>
<td>CY, EL, HR, LI, UK (E&amp;NI, Wales, Scotland)</td>
</tr>
</tbody>
</table>

Source: 2015 Cedefop European Inventory for Validation

The majority of countries have multiple frameworks covering different education sectors (VET, school and higher education Acts that enable formal education and training institutions to recognise learning outcomes acquired in non-formal and informal learning).

The disadvantage of a legal framework is that the system is less agile in reacting to changes, in comparison to countries with multiple frameworks. Multiple frameworks that lead to multiple validation processes can also make it difficult for the public to understand, or to mainstream processes, for example regarding quality assurance.

Challenges remain at the level of implementation, which is uneven across countries. A literature review by Stenlund (2010) on the assessment of prior learning in higher education concludes that there is a need for greater consistency in the procedures of validation both in and among universities and education programmes, as some claimants are disadvantaged, depending on what university or faculty they choose, and on the instruments employed in the validation process. Clear legal frameworks that are rigorously implemented could help in this context. In-depth discussions on financial sustainability are often absent in existing legal frameworks.
### Table 1.3: Countries with VNFIL arrangements by sector of education and for which centralised data are available

<table>
<thead>
<tr>
<th>Sector</th>
<th>Number of countries with VNFIL arrangements in place</th>
<th>Number of countries with VNFIL procedures in place and for which centralised data are available</th>
</tr>
</thead>
<tbody>
<tr>
<td>General education (secondary level)</td>
<td>16</td>
<td>8 (BE-FL, DE, DK, IS, NL, NO, PL, PT)</td>
</tr>
<tr>
<td>Initial VET</td>
<td>27</td>
<td>12 (AT, BE-FL, CH, DE, FI, FR, LI, LU, LV, NO, PL, PT)</td>
</tr>
<tr>
<td>Continuing VET</td>
<td>25</td>
<td>8 (BE-FR, CZ, DK, FI, LV, MT (childcare sector only), NO, ES)</td>
</tr>
<tr>
<td>Adult education</td>
<td>15</td>
<td>5 (AT, BE-FR, IE, IS, LV)</td>
</tr>
<tr>
<td>Higher education</td>
<td>23</td>
<td>7 (AT, FR, BE-FR, LV, NO, UK-England and Northern Ireland, UK-Wales)</td>
</tr>
</tbody>
</table>

Source: Cedefop, 2016

### 2. Validity and reliability of assessments for VNFIL

The effectiveness of VNFIL assessments depends upon the validity and reliability of the information gathered in the assessment process. Validity refers to the degree to which the assessment measures what it is intended to measure (i.e. occupational standards or standards and learning outcomes in national systems and education and training policies). In addition, validity depends on the meaningfulness, usefulness and fairness of the assessment (Kane, 2006; Messick, 1989 – cited in Lane 2010). The purpose, design and uses of the assessment should be specified to ensure that evidence gathered is appropriate. In addition, assessments should be free of bias, and should ensure equitable treatment during the testing process and in supporting opportunities to learn (Condelli and Baker, 2002). Intended and unintended consequences of the uses of assessments should also be considered (Messick 1994).

Reliability refers to the consistency and stability of assessment results across populations. Reliability helps to ensure that certificates, no matter where they are issued, will have equal value for employers and/or educational institutions. Variations in rating may occur when there are no agreed-upon criteria or benchmarks, when there are differences in assessors’ attitudes or bias (Davidson, Howell and Hoekema, 2000 cited in Jonsson and Svingby, 2007). Poorly designed or inconsistent types of tasks may also contribute to a lack of reliability (Shavelson, Gaoand Baxter, 1996, cited in Jonsson and Svingby, 2007; Hathcoat and Penn, 2012).

Standardised assessments, which are carefully administered to ensure the same conditions for all test takers and scored in a consistent manner, are considered as being more reliable. They are also more cost effective than alternative assessments.

Assessments should also be usable. That is, they should be easy to administer and to interpret (Abu-Alhija, 2007). When used formatively, the results should point to next steps for career development and/or academic work. When used summatively, the assessment should be clearly aligned with learning outcomes and standards, to ensure transparency of qualifications.
Learning outcomes: “sets of knowledge, skills and/or competences an individual has acquired and/or is able to demonstrate after completion of a learning process, either formal, non-formal or informal” (Cedefop, 2014, pp. 164-165).

Competence: the “ability to apply learning outcomes adequately in a defined context (education, work, personal or professional development)” (Cedefop, 2014, p. 47).

Assessment of learning outcomes: the “process of appraising knowledge, know-how, skills and/or competences of an individual against predefined criteria (learning expectations, measurement of learning outcomes)” (Cedefop Terminology 2014).

Formative assessment: “a two-way reflective process between a teacher/assessor and learner to promote learning” (Cedefop, 2015, European Guidelines VNFIL, p. 74).

Summative assessment: “The process of assessing (or evaluating) a learner’s achievement of specific knowledge, skills and competence at a particular time” (Cedefop, 2015, European Guidelines VNFIL, p. 77).

Standardised assessments: “assessments that are developed, administered, scored and graded according to uniform procedures designed to ensure consistent outcomes that can be meaningfully compared across a population” (EC, 2012, p. 31).

Alternative assessments: “based on the use of ‘innovative’ methods of assessment ... portfolios, self and peer assessment and simulations – amongst other methods – as opposed to traditional multiple choice tests and essay writing” (Cedefop, ICF International, 2014)

Recognition of learning outcomes: the “process of granting official status to knowledge, skills and competences either through:
- validation of non-formal & informal learning
- grant of equivalence, credit unions or waivers
- award of qualifications (certificates, diploma or titles) and/or
Social recognition: acknowledgment of value of knowledge, skills and/or competences by economic and social stakeholders” (Cedefop Terminology 2014)

Validation of learning outcomes: “confirmation by a competent body that learning outcomes acquired by an individual in a formal, non-formal or informal setting have been assessed against predefined criteria and are compliant with the requirements of a validation standard. Validation typically leads to certification
Or
Process of confirmation by an authorised body that an individual had acquired learning outcomes measured against a relevant standard
Validation consists of four distinct phases:
- identification through dialogue of particular experiences of an individual
- documentation to make visible the individual’s experiences
- formal assessment of these experiences; and
- certification of the results of the assessment which may lead to partial or full qualification” (Cedefop Terminology 2014).

Formal learning: “learning that occurs in an organised and structured environment (such as in an education or training institution or on the job) and is explicitly designated as learning (in terms of
objectives, time or resources, formal learning is intentional from the learner’s point of view. It typically leads to certification” (Cedefop Terminology 2014).

Informal learning: “learning resulting from daily activities related to work, family or leisure. It is not organised or structured in terms of objectives, time or learning support. Informal learning is in most cases unintentional from the learner’s perspective” (Cedefop Terminology 2014).

Non-formal learning: “learning embedded in planned activities not explicitly designated as learning (in terms of learning objectives, learning time or learning support). Non-formal learning is intentional from the learner’s point of view” (Cedefop Terminology 2014).

Qualifications Framework: “instrument for development and classification of qualifications (at national or sectoral levels) according to a set of criteria (using descriptors) applicable to specified levels of learning outcomes”

Or
“instrument for classification of qualifications according to a set of criteria for specified levels of learning achieved, which aims to integrate and coordinate qualifications subsystems and improve transparency, access, progression and quality of qualifications in relation to the labour market and civil society” (Cedefop Terminology 2014)

Qualification system: “all activities related to the recognition of learning outcomes and other mechanisms that link education and training to the labour market and civil society. These activities include:
- definition of qualification policy, training design and implementation, institutional arrangements, funding, quality assurance
- assessment and certification of learning outcomes” (Cedefop Terminology 2014)

European Qualifications Framework (EQF): “reference tool for describing and comparing qualification levels in qualifications systems developed at national, international or sectoral levels” (Cedefop Terminology 2014).

As noted at the beginning of this report, there is concern that these “traditional” assessments may be inappropriate for non-traditional learners (e.g. immigrants, refugees, individuals with low qualifications, individuals with disabilities). Tsekoura and Giannakopoulou (2017), in a companion report developed for the InnoVal project, note that barriers for individuals in disadvantaged groups include low language and/or literacy skills, poor digital skills (necessary for some forms of assessment, such as the digital badge), and/or the stress associated with formal examinations. Individuals seeking accreditation may also object to processes that do not directly lead to accreditation (e.g. assessments that are used for formative purposes, or a certification which is not recognised by public or private sector employers).

In the InnoVal online consultation conducted for this project, all respondents reported that they work with individuals who would likely benefit from assessments tailored for diverse needs. Of these respondents, 48.1% serve early school leavers, 45.6% serve immigrant learners, 41.8% serve refugee learners and 41.8% serve individuals with disabilities. They serve a mix of youth – that is individuals between the ages of 16 and 19 (85.5%) and adults between 30 and 65 years of age (74.7%).

While these respondents view current approaches to VNFIL in a mostly positive light, a significant percentage of respondents to the online consultation indicate that there is room for improvement.
For many providers, however, the advantages of standardised assessments appear to outweigh the disadvantages. Indeed, a majority of stakeholders responding to the InnoVal online consultation on barriers and enablers to VNFIL agreed that standardisation of assessments is essential to ensure validity and reliability of results – with 52.4% strongly agreeing, and 37.8% partially agreeing.

Several respondents qualified their choices, noting that while standardisation is important for mobility in the labour market and for educational institutions, assessments must be fit for purpose. For example, an employment counselor in Belgium indicated that while standards are necessary, a welder should not be assessed through a written examination. Indeed, while it is possible to develop written assessments that ask candidates to describe a specific process important for the certificate, a simulation or live observation will provide much better evidence of the candidates’ performance level.

This question also drew a number of additional respondent comments, among them:

- “The validation process must take into account that each individual is unique. That may challenge the
level of standardisation.”
- “Non-formal/informal learning by its very nature shouldn’t be completely standardised as it won’t fit in a formal accreditation system”.
- “We need to have multiple forms of assessment available. There cannot be one standard assessment method.”

(The full set of comments is available at the InnoVal publication: “Public Consultation on Barriers and Enablers: Analysis of Results”)

As the stakeholders recognise, standardised approaches cannot capture all aspects of competences. Other analysts have also noted that while standardised assessments that are based on multiple-choice questions may be used to assess higher order knowledge, they cannot be used to measure competences such as the capacity to develop an argument. Testing methodologies that treat tasks as discrete questions cannot easily capture complex performances, interactions or processes. If multiple-choice assessments are poorly designed, they may also be prone to measurement error (e.g. students may misinterpret questions or may make random guesses) (Looney, 2011).

Finally traditional assessments do not easily capture learners’ creativity, or ability to think beyond prescribed parameters or learning outcomes (Tsekoura and Giannakopoulou, 2017; Cedefop, 2016), or differences in cultural background experience of immigrant learners (Klayer and Arend Odé, 2003).

### ADVANTAGES AND DISADVANTAGES OF DIFFERENT ASSESSMENT FORMATS

| **Tests and examinations:** | Include multiple-choice assessments and written assessments (e.g. essays). Multiple-choice assessments provide reliable data on student performance, as assessment are machine-scored, and are therefore less expensive to administer. Well-designed multiple-choice questions may be used to assess higher-order knowledge. They cannot, however, measure skills such as the capacity to develop an argument. Poorly designed multiple-choice assessments are also prone to measurement error (e.g. students may misinterpret questions or may make random guesses). Written assessments provide opportunities for candidates to demonstrate complex problem solving skills and higher-order knowledge.
|
| **Portfolios:** | Candidates gather a range of documents, videos or other media that illustrate competences as applied in a range of contexts and circumstances. The validity and reliability of the portfolio depends on the completeness of the dossier, the clarity of criteria and expected outcomes, as well as the quality of guidance.
|
| **Declarative methods:** might be included in a portfolio. The candidate is asked to record competences gained through non-formal and informal learning. They may be used as a screening tool to determine whether the candidate is ready for a full validation procedure, or to identify needed competences. The validity and reliability of declarative methods depends on the clarity of criteria and expected outcomes, as well as the quality of guidance.
| **Interviews:** Candidates may provide information on competences through dialogue with a counselor or
assessor. This dialogue format allows the assessor to pose follow-up questions in order to clarify and avoid misunderstandings. The quality of the information garnered in the interview process is affected by the interview environment, the questioning skills of the assessor and communication skills of assessor and candidate.

Interviews may be particularly challenging for migrant or refugee learners with limited skills in the language of the host country.

**Observations**: Candidates are observed performing everyday tasks relevant to the area assessed. The observations are typically carried out by a trainer/instructor and/or individual qualified professional. The candidate may be asked to describe what they are doing and why. He/she may also be asked to assess his/her own performance. The validity of observation methods is considered as high, however the method is time-consuming. Moreover, the competences observed, which are performed in a specific context, may not be transferable to other contexts.

**Simulations**: In situations where it is not possible to set up a real-life observation, a simulation of a real-life scenario may be created. Assessments, typically by a jury or panel, are based on defined standards and criteria (alternatively, the candidate may participate in an ICT-based simulation).

Simulations require careful preparation and research for development of appropriate scenarios. They are thus costly and time-consuming. Simulations are considered as having high reliability and validity. However, as with observations, it may not be possible to draw broad conclusions on competences performed in a specific context.

Computer-based performance assessments may potentially assess more complex performances through simulation, interactivity, collaboration and constructed response formats. Increasingly sophisticated ICT programmes that score “open-ended performances” (including simulations) may address concerns regarding reliability of human-scored assessments, and validity of multiple-choice assessments that do not effectively measure higher-order skills (OECD 2013).

As mentioned above, mixed approaches, which bring together standardised and non-standardised elements, are increasingly used (Souto-Otero and Villaba-García, 2014). This is important as no single assessment can provide enough information to capture an individual’s competences. Moreover, different assessment methods provide different ways of measuring competences, and may therefore provide different results. A combination of standardised and alternative assessments, gathered over time, will create a more well-rounded picture of an individual’s competences.
3. Priorities for development

While there has been significant progress in developing effective assessments for VNFIL, there is room for improvement. For example, stakeholders responding to the InnoVal online consultation identified the three most significant technical barriers to implementing alternative approaches to VNFIL as the difficulty of training human assessors to ensure reliability, the cost of administering alternative approaches, and insufficient development of ICT-based measurement technologies.

The following sections explore needs for development in the areas of:

- Capacity building
- Improving the quality of alternative assessments
- Matching assessments to needs
- Cost considerations
- Alignment with National Qualifications Frameworks
- Stakeholder engagement

Below, we discuss these challenges in more detail, and highlight approaches to addressing barriers within current technical and financial constraints.

Capacity Building

Assessors (or counselors) play a necessary and vital role in the VNFIL process. They may use assessments formatively – to identify competences and provide guidance on next steps for career and/or academic development, or summatively – to make decisions regarding certification, employment or admissions to educational programmes. In both cases, appropriate training can help build capacity.

Assessors/counselors need a range of social-emotional as well as technical competences. For example, the National Knowledge Centre for Validation of Prior Learning in Denmark has noted that the counselors should:

- Treat the candidate with openness and respect
- Provide complete information and counseling on options and the process of validation
- Use assessment tools to interact with the candidate and clarify meanings, expectations and motives
- Counsel the candidate on options to further develop competences or on employment opportunities
- Ensure transparency of the process
- Tailor the approach to the individual candidate’s needs

(Aargard, K., Unpublished draft paper prepared for the Bertelsmann Foundation, 2014)

In terms of technical requirements, the task of ensuring reliability of summative assessments is of most concern for stakeholders of VNFIL. Caldwell and colleagues (2003) have found, however, that effective training can improve the reliability of scores on performance-based assessments.

Collective work among jury members may also build the assessment capacity of the individual members. Sandrine (2012) notes that coordination meetings of jury members involved in validation of prior learning as help to foster their collective learning and improve the assessment outcome. In the context of the InnoVal case studies, it will be important to explore how assessors have been trained, and the extent to which colleagues
using a particular assessment tool agree with each others’ scoring judgments, and why or why not.

Cedefop’s 2014 inventory notes that a number of countries support continuous professional development of staff involved in validation process across all relevant sectors. Guidance counsellors also need to have a common understanding of the way a validation method should be carried out, but training is scarce. More commonly, countries support counsellors through guidelines and tools (Cedefop European Inventory VNFIL 2014). It is more common for validation staff to have a right to training in initial and continuing vocational education and training and general education, than in adult education or higher education (Cedefop European Inventory VNFIL, 2016).

Examples of training provision and qualifications to support the professional development validation professionals

- In Luxembourg, training is systematically provided to guidance practitioners that support candidates and members of validation committees.
- In Romania, training for validation practitioners is provided by the validation centres but not in a regular basis.
- In Slovakia, assessors must have passed an examination for practicing as adult educators/trainers (lector).
- In Bulgaria, specialized training courses have been developed for validation practitioners involved in the “System for validation of non-formal acquired knowledge, skills and competences” project.
- In the UK-Scotland, a few organizations in the public sector have introduced training for staff (and stakeholders), also making available resources materials and workshops opportunities.
- In Norway, country authorities provide training to assessors in primary and upper secondary education on an annual basis and inexperienced assessors are also given mentoring support.
- In Estonia, there have been many training opportunities in the HE sector due to the extensive funding (ESF programme). Training courses are offered to assessors, councillors/advisers and applicants. During the period 2008-2012, 973 people participated in RPL assessor training and 242 in RPL councilor training.
- In Iceland, the ETSC training course for assessors was developed for the Ministry of Education, Science and Culture (MESC) in 2008 and has then offered training for project managers, assessors, and counsellors/advisers. It was a two-day course where the concept and practices are reviewed and discussed.
- In the UK, the Agored Cymru Level 3 Award in Recognition of Prior Learning (RPL) is intended for practitioners working at any level in the education and training sector across the UK. The qualification, which is currently being introduced into the QCF, is made up of three units: the theory of RPL; formative RPL and summative RPL. It is hoped that the introduction of the qualification should lead to greater degree of consistency in the application of RPL.
- In Ireland, an example of training provision is the Cork Institute of Technology and the Dublin Institute of Technology, which provide a range of training opportunities for training to staff who have responsibility for the development and management of any aspect of RPL. This includes an internal staff website; briefing sessions; workshops; consultations and formal 5 ECTS credit training course at Masters level for policy and academic staff.
- In Switzerland, the Certificate and Diploma of Advanced Studies in Validation is organized by the Swiss Federal Institute for Vocational Education and Training (SFIVET), the academic institution in charge of qualifying and certifying VET teachers and trainers. This modular qualification path includes a specific module aimed at specializing professionals in advising, assessing and managing the validation procedures. Every module delivers 5 ECTS.
- As noted above, in time, validation professionals in Malta will be required to hold a qualification in the specific area and a specifically-developed qualification (for assessors) at the MQF/EQF Level 4.
- In Finland, assessors must have the diploma of Specialist in Competence Based Qualifications: a 25 credit compulsory training programme.

Source: Cedefop European Inventory VNFIL 2014
Improving the quality of alternative assessment

A number of measurement experts have devoted attention to strengthening the quality of alternative assessments. They have piloted new design frameworks that allow assessors to capture and rate complex performances, and to ensure validity and reliability. These pilots and programmes include simulation-, scenario-, interview- or portfolio-based assessments for both live and virtual performances. The research is typically focused on school-level assessment or on training in fields such as medicine, piloting or disaster triage, but the findings are also relevant to VNFIL.

Designing assessments

Tools for performance-based assessment which are subject to rigorous design and piloting processes have high levels of validity and reliability. Once developed, assessment templates allow for easy adaptation and realise cost efficiencies.

This section highlights that the design of performance-based assessments and training of assessors can improve the validity and reliability of assessments. The addition of standardised questions can also increase reliability. There are several possible approaches to ensuring that the assessment will measure what it is intended to (i.e., that it is valid), and that scores will be consistent across raters (i.e., that it is reliable).

Evidence-centred design

Mislevy (2011) describes a four-step evidence-centred design (ECD) process for simulation-based assessments (applicable to both live and ICT-based simulations). The first step is to define the target population, the purposes of the assessment, the content and skills to be covered, and the number and types of tasks needed to assess a particular skill. Tasks chosen may either provide evidence for a larger domain of interest or a more targeted skill (Lane (2010) refers to the latter approach as the “merit badge approach). Mislevy (2011) recommends that interdisciplinary teams involving experts in the domain to be tested, as well as measurement experts (and for ICT-based simulations, software designers) work together as an interdisciplinary team. Domain experts help to identify knowledge and skills needed, types of situations in which they are used, and the different problem-solving strategies typically used by novices versus experts.

Following this, measurement experts develop models for tasks and processes with “evidentiary value” (the Conceptual Assessment Framework, of CAF). They then author tasks and scoring details (for a rubric or for automated scoring in an online assessment). In the fourth stage of this process, assessments are piloted and adapted. Measurement experts also make decisions about the number of tasks to include, time spent on testing, and whether some should be standardised.

To test the quality of the design, test developers review tasks to be sure they are clear, are directly relevant to the domain being assessed and learning outcomes measured, are likely to elicit responses which can be scored reliably, and avoid potential bias. Another approach is to ask assessees to “think aloud” to describe their process, or describe it retrospectively. This will help test developers to ensure the problems stated are clear and not subject to misconception or errors in understanding. For assessments with higher stakes, larger scale field tests and statistical evidence of validity and reliability may also be needed (Condelli and Baker, 2002).

After pilots and field tests are complete, assessment developers create the final assessment tool and develop detailed scoring guides for assessors, along with a standardised administration procedure (Condelli and Baker, 2002). These steps all help to strengthen the validity and reliability of a performance-based assessment.
Scoring rubrics

Rubrics are scoring tools that support qualitative or quantitative scoring of performances. They identify elements of performance and set out criteria and standards at different levels to support transparent, fair and internally consistent scoring.

Rubrics may be holistic or analytical. With the holistic approach, the assessor rates the overall quality of the performance (usually used with large-scale assessments). With the analytical approach, the assessor rates the strengths and weaknesses of different elements of the performance and identifies learning needs (Condelli and Baker, 2002; Jonsson and Svingby, 2007).

Condelli and Baker (2002) recommend that rubrics clearly distinguish score levels, with approximately equal increments between each level. The rubric should also use descriptive vocabulary and avoid terms that are vague or judgmental. Tighter restrictions may increase reliability, but the rubric should balance reliability with validity for authentic forms of assessment (Jonsson and Svingby, 2007).

The process designing rubrics to assess complex knowledge and processes involves an up-front investment. However, templates for performance tasks that assess the same cognitive process and competences may be developed, and may facilitate the development of other performance-based assessments. These templates also improve the generalisability of scores and may be used for live or computer-based simulation tasks. (Lane, 2010). There are thus efficiencies as initial designs are adapted.

Additional strategies to strengthen validity and reliability

Other strategies may be to include more tasks on the assessment or to lengthen the time spent. For example, Hathcoat and Penn (2012) found that more tasks on assessment and standardisation of some tasks increase the reliability of results. In another study focused on assessment of medical trainees, Van Der Vleuten and Schuwirth (2005) found that reliability was more closely tied to the time spent in the testing rather than the method. Reliability of results after one hour of testing using a range of methods (including multiple-choice tests, oral examinations, case examinations, practice video assessment, mini-clinical exercises, incognito standardised patients) was low, but increased substantially after eight hours of testing. For performances judged by juries, a unifying theme (appropriate for the field being assessed) that runs through all domains of performance being assessed, may also improve reliability. Jury members may more easily balance assessment of complex factors involved in performance-based assessment (Crossley et al., 2011). Abedi and Lord (2001) found that assessments that minimise linguistic demands (e.g. with graphic displays) have led to better outcomes for non-native speakers as well as for individuals with poor literacy or linguistic skills.

For both standardised and performance-based assessments, more research on how we learn and how we measure that learning is needed. In addition, there is a need for much more research on measurement of transversal skills (also sometimes referred to as “soft skills”). There is no general consensus on the best approaches to assessing transversal skills. Nor do qualifications frameworks include standards and learning outcomes for transversal skills such as communications skills, analytical and problem solving skills, ability to adapt to and act in new situations, decision-making skills, teamwork skills, and planning and organisational skills (Eurobarometer, 2010) – and indeed, it would be very difficult to set standards for such traits. Nevertheless, it will be important to provide a way to recognise and value these traits.
Training human assessors

High-quality tools for performance-based assessment tools are essential, but assessors also need careful training. Condelli and Baker (2002) recommend that this include training on how to use the rubrics, exemplars showing how experts have scored responses, and opportunities to practice scoring. Quality checks by individuals who are experts in content as well as scoring may randomly check the quality of assessors’ scoring. Lane and Stone (2006) recommend that assessors may need to spend several days together in order to discuss their interpretation of scoring criteria, and the extent to which they are stricter or more lenient in scoring, and any potential sources of bias. Such training can help to strengthen reliability, and to the extent that it is possible to maintain stable scoring teams, may be worth the investment.

Matching assessments to needs

Above, we noted that variety of assessment methodologies and sources of data standardised assessments as well as performance-based assessments, such as portfolios, interviews, simulations, workplace observations, etc. will provide a more accurate picture of the candidate’s performance. Typically, a mix of standardised assessments (with higher levels of reliability) and human-rated performance based assessments are part of the mix.

However, assessors also need to make decisions on how to weight different assessments, how to choose among complementary approaches, and how to interpret data gathered through different methods, whether for formative or for summative purposes (Baker, 2003). In the case that the candidate will eventually need to take a traditional, standardised examination, the assessor will also need to decide how to best prepare the individual. Indeed, the majority of respondents to the InnoVal stakeholder consultation only partially agreed with the statement that “standardised assessments have a negative impact on candidates’ decisions to pursue VNFIL.”

Table 3.1: Standardised assessments have a negative impact on candidates’ decisions to pursue VNFIL

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Partially agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
<th>Don’t know</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher education</td>
<td>3.3%</td>
<td>53.3%</td>
<td>20% ----</td>
<td>20%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Adult education</td>
<td>52.9%</td>
<td>29.4%</td>
<td>20%</td>
<td>10%</td>
<td>5.9%</td>
<td>11.7%</td>
</tr>
<tr>
<td>Occupational training</td>
<td>12.5%</td>
<td>50%</td>
<td>37.5%</td>
<td>10%</td>
<td>20%</td>
<td>60%</td>
</tr>
<tr>
<td>Public authorities</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Employment counseling</td>
<td>40%</td>
<td>60%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Non-formal &amp; informal learning</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Source: InnoVal Online Consultation Analysis, 2017
There were a range of open responses to this question. One respondent noted with proper support from the institution organising the VNFIL process, candidates can succeed. Another respondent noted that motivation for certification can inspire candidates to overcome negative associations of traditional, standardised assessments.

For formative assessments, assessors are more likely to have leeway in the type of assessment method used, and may choose as they see fit, based on the candidates’ profile, or the type of competence to be assessed. Galanis and colleagues (2016) propose that learners may bolster each others’ informal learning through use of a flexible online peer-driven evaluation and validation framework. This type of social collaboration for informal learning allows users’ to gather a range of suggestions and learn from peers’ experiences.

For summative assessments, for example, in the case of certification, in most countries there is much less room for choice of method, either on the part of the assessor or the candidate. Norway, highlighted in the box below, is an exception as there is no prescribed or standardised method for validation.

**NORWAY**

Providers at the local level are not required to follow a standardised procedure for validation. There is a four-stage process that may be applied for any level of education and training:

1. Information and guidance;
2. Identification and systematization of competences;
3. Assessment; and
4. Documentation.

Candidates do not take “traditional” exams. Rather, the provider decides on the method that is most relevant for the individual candidate. Methods include documentation (e.g. through a portfolio), a short written test, or an interview/dialogue to describe competences. The method is to be tailored to the individual needs, but the main aim is to ensure that there is a clear view of the candidate’s competences. With this in mind new guidelines state that portfolios should be combined with an interview or dialogue in order to describe competences more fully.


An important part of the InnoVal case studies will be to understand how individual counselors/assessors decide which assessment tools are appropriate, for which learners, and when. Immigrant learners may also need more targeted and tailored support in order to make their own informed judgments about the relevance of their prior experiences, as well as for navigating new and very different systems (Klayer and Arend Odé, 2003).
Cost considerations

A significant number of stakeholders responding to the InnoVal online consultation also expressed concerns regarding the cost of administering alternative approaches. Decisions on trade-offs between quality and cost (or feasibility) are perennial and unavoidable. As noted by Stecher et al (1997),

…it is usually not possible to maximize both quality and feasibility, so vocational educators must strike a balance between them. As assessment becomes more authentic, it also becomes more expensive to develop, to administer, and to score. In addition, greater quality usually involves greater costs and greater commitment of time. This is no simple formula for balancing these factors. (p. 35)

Potentially, new and innovative approaches can help lessen the cost of developing and administering assessments, although some up-front investments at the system level would be needed. However, for the purposes of InnoVal, the main concern is how decision makers at the provider level are balancing these costs, and the ability to meet learner/candidate needs as effectively as possible.

Alignment with National Qualifications Frameworks

Respondents to the InnoVal stakeholder consultation had a range of views on the degree to which VNFIL assessment should be aligned with National Qualifications frameworks. Overall, 35.8% of respondents strongly agree with this statement, and 37.8% partially agree. There is variation across the sectors, however, as illustrated in the table below.

Table 3.2: VNFIL assessment should be aligned with National Qualifications frameworks

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Partially agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
<th>Don’t know</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher education</td>
<td>33.3%</td>
<td>40%</td>
<td>26.7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult education</td>
<td>29.4%</td>
<td>41.2%</td>
<td>5.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupational training</td>
<td>62.5%</td>
<td>25%</td>
<td>12.5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public authorities</td>
<td>40%</td>
<td>20%</td>
<td>40%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment counseling</td>
<td>40%</td>
<td>40%</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-formal &amp; informal learning</td>
<td>60%</td>
<td>20%</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: InnoVal Online Consultation Analysis, 2017

It should be noted that some commentators have cautioned against overly tight alignment and narrow definitions of learning outcomes in qualifications frameworks. Learning outcomes that are closely tied to specific job skills may prevent individuals from advancing to higher levels of learning or better job opportunities. These approaches also underemphasise skills such as critical thinking and problem solving. Broader and more holistic approaches support skill transfer and more open pathways for lifelong learning (Allais, 2011; Young, 2007).
Stakeholder engagement

Stakeholder engagement is difficult to gauge. It ranges from awareness to engagement in the design or policies, tools and processes. The majority of respondents to the InnoVal consultation only partially agree that stakeholders are sufficiently involved in the development of assessment approaches and tools for VNFIL.

Table 3.3: Stakeholders are sufficiently involved

<table>
<thead>
<tr>
<th>Category</th>
<th>Strongly agree</th>
<th>Partially agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
<th>Don’t know</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher education</td>
<td>16.7%</td>
<td>23.3%</td>
<td>26.7%</td>
<td>6.7%</td>
<td>10%</td>
<td>13.3%</td>
</tr>
<tr>
<td>Adult education</td>
<td>17.7%</td>
<td>29.4%</td>
<td>29.4%</td>
<td>11.8%</td>
<td>11.8%</td>
<td></td>
</tr>
<tr>
<td>Occupational training</td>
<td>62.5%</td>
<td>25%</td>
<td></td>
<td></td>
<td></td>
<td>12.5%</td>
</tr>
<tr>
<td>Public authorities</td>
<td>80%</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment counseling</td>
<td>20%</td>
<td>60%</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-formal &amp; informal learning</td>
<td>40%</td>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td>20%</td>
</tr>
</tbody>
</table>

Source: InnoVal Online Consultation Analysis, 2017

Aggarwal (2015) emphasises the importance of effective stakeholder engagement (assessees, employers, employees, community, government and education and training providers) to improve the quality and acceptance of alternative VNFIL assessment. This broad engagement also helps to build broader awareness of processes and impacts.
4. A Summary of Needs for Innovation in VNFIL

The above analysis points to several needs for further development and innovation of alternative assessments of VNFIL. These include:

**Investments in assessment design (standardised and performance-based)**

- Content and measurement experts need to be involved in the design process
- Assessments need to be adapted for disadvantaged groups (literacy and language and cultural needs, disability, and so on)
- Assessment design provides information needed for next steps in guidance and counseling [formative purposes])
- Guidance on how to assess transversal skills
- R&D in computer-based assessments

**Effective training of human assessors**

- Assessors/counselors learn how to develop an appropriate mix of standardised and ‘authentic assessments’
- Assessors/counselors ensure assessments are appropriate for candidates’ needs/help yield maximum information on each candidates’ competences], support take-up
- Assessors have the opportunity to develop a shared understanding of standards and scoring criteria

**Benefit-cost analysis of investment in performance-based assessments**

- Take-up levels for individuals in disadvantaged groups
- Outcomes (further education, certification, employment)
- Equity

Human assessors will need to know when to use what type of assessment and for what purpose, and the strengths and weaknesses of each different type of assessment. They also need an understanding of how to develop a more holistic view of learners’ competences, and how to weight different assessments.

Systems will also need to make decisions on tradeoffs across standardised assessments that ensure greater reliability, and alternative assessments that measure complex skills.

Finally, greater stakeholder involvement will be vital for ensuring quality, transparency and acceptance of alternative assessments.
REFERENCES

Aargard, K., “The Danish system on validation of non-formal and informal learning”, unpublished draft paper prepared for the Bertelsmann Foundation, 2014


Crossley et al.(2011), “Good questions, good answers: construct alignment improves the performance of


Uras, F. (2017), Consultation on Barriers and Enablers: Analysis of consultation results, InnoVal Project

PROJECT PARTNERS

- LIFELONG LEARNING PLATFORM (Belgium)
- DAFNI KEK (Greece)
- INSTITUT EUROPEEN D’ÉDUCATION ET DE POLITIQUE SOCIALE - EIESP (France)
- ANESPO (Portugal)
- UC LEUVEN-LIMBURG (Belgium)
- EUROPEAN UNIVERSITIES CONTINUING EDUCATION NETWORK - EUCEN (Belgium)
- Associate Partner: BERTELSMANN STIFTUNG (Germany)