Assessing adults' quantitative skills – the INULIS project

John O' Donoghue

Department of Mathematics & Statistics,
University of Limerick

John.ODonoghue@ul.ie

Henk Van Der Kooij

Freudenthal Institute
GE Utreacht

henk@fi.uu.nl

Improving Numerical Literacy Skills (INULIS) project is a Leonardo Da Vinci vocational training improvement project coordinated by the Greek technical company Cyberce with the University of Limerick in Ireland, the Freudenthal Institute of the University of Utrecht in Netherlands, the Traders Association of Thessaloniki in Greece and Consulo, a communications company in Cyprus as partners.

The main aim of the INULIS project is to develop a practical on-line assessment instrument that can be used by HR personnel to assess a range of 'quantitative knowledge and skills' suitable for workers to gain, maintain and advance in employment in Small and Medium-sized Industries (SMEs) and commercially as small traders. Such tests will be constructed from an electronic database containing 2400 pre-generated test items. Users of the database will have access to the methodology for constructing on-line tests, and scoring and interpreting these tests.
In this paper the authors address issues such as project aims and rationale, item production and types, test construction and theoretical underpinnings.

INULIS Project rationale

There is no universally accepted view as to what constitutes 'quantitative knowledge and skills' but there are many variants of the concept in the research literature such as 'numeracy' (Cockcroft, 1982; Adult Literacy and Life Skills (ALLS) Survey, 2000); 'quantitative literacy' (International Adult Literacy Study (IALS), OECD, 1995); 'mathematical literacy' (Programme for International Assessment (PISA), OECD, 2003). The INULIS partners decided not to problematize the concept of 'quantitative knowledge and skills' but rather to proceed on the basis of a suitable definition or characterisation obtained from the relevant research literature.

There are no definitive studies that give a clear direction to follow but some approaches have more to offer than others. The INULIS partners recognised a clear choice between studies that offer lists of basic mathematical skills, mainly computational skills, and others that attempted to define a mathematical competence. There is sufficient common understanding among INULIS partners of the meaning attaching to 'quantitative knowledge and skills' in terms of school mathematics, the numeracy/quantitative literacy debate internationally and the requirements of workers, to accept that there is sufficient congruence, resonance or overlap with PISA's 'mathematical literacy' concept to proceed with confidence.

The strategy adopted by the INULIS project team was to work forward to its goal of a ‘fit for purpose test’ by systematically following the PISA approach where that was practicable and feasible in terms of INULIS resources and otherwise adapt the PISA approach when necessary. PISA is used as an exemplar of ‘good practice’ and a methodological guide.

This strategy offered important immediate advantages:

- A common understanding of the meaning of ‘quantitative knowledge and skills’
- Clear guidelines on appropriate mathematical competence(s) to meet the needs of individual’s current and future life
- An assessment framework and methodology, and
- An international benchmark for good practice.

However, a very challenging mathematical education task still remained: to produce a ‘fit for purpose’ assessment instrument that could be constructed on-line and delivered on-line from a specially developed database of 2400 test items.

INULIS Database

A large database of well-constructed test items is absolutely essential for the successful attainment of the INULIS project goals. It is not surprising, therefore, that a major effort and a significant proportion of project resources were dedicated to the development of the database. This section deals with the nature and production of test items and quality control issues. It does not deal with the technical specification of the database.

Item development

The production of test items for the database was the responsibility of the Irish, Dutch and Greek partners. The workload was distributed as follows (Table 1).

Table 1 Distribution of workload

Partner	Number of items
CAMET(Ireland), University of Limerick, Ireland	840
Freudenthal Institute University of Utrecht Netherlands	840
Cyberce Athens Greece	720

The development teams, between them, included a wide range of expertise in mathematics, school mathematics, mathematics education, numeracy and assessment.

The aim of the item development teams was to fill the database with mainly computer-scorable items. Each partner was responsible for developing items over the whole range of themes, competencies and contexts. In contrast to PISA (a pencil and paper test), more variations in types of items was possible because of the technological features a computer can offer. Computer-scorable items include multiple-choice and other constructions (numbers, pictures, objects) that are recognised by software, and also make possible the use of hotspots etc. See appendix A for issues related to quality control.

The INULIS database contains a number of item types as described in Table 2.

Table 2 Question Types for INULIS Application

Code	Question Type	Notes
MC	Multiple Choice (4 answers)	4 alternative answers provided in the form of text or images. The user has to select ONLY one GUI: Radio buttons
SDL	Scroll Down List	A few alternative answers in the form of text. The user has to select ONLY one GUI: Scroll down list
B	Boolean	2 alternative answers in the form of text. The user has to select ONLY one (YES/ NO, True/ False) GUI: Radio buttons
CB	Complex Boolean	A few answers provided in the form of text or images. The user has to select for each one if the answer is YES/ NO, True/ False, Right/ Wrong, etc. GUI: Radio Buttons (for each answer)
	Short Response	
SRS	Single Answer	The user provides 1 answer in the format of a number and in some cases he/ she has also to select from a list the unit of measurement GUI - Free Text (for number, i.e. 11.121,32) - Scroll down list (for the unit of measurement) Note: The number may have been predefined as ranges of right numbers by the test tutor (i.e. 300 – 500)
SRG	Grid-based Answer	The user provides >1 answers in the format of a number and in some cases he/ she has also to select from a list the units of measurement per answer GUI (in the form of tables) - Free Text (for numbers, i.e. 11.121,32) - Scroll down lists (for the units of measurement) Notes: - The numbers may have been predefined as ranges of right numbers by the test tutor (i.e. 300 – 500) - The case that the user has to provide "Min, Max" values is considered as a Grid-based question (table = 1,2)
DD	Drag n' Drop	The user can drag and drop >1 objects (i.e. images) in the right order GUI: Object areas Note: The case that the user has to "Match" objects with places is considered as a special case of Drag and Drop question type
HS	Hot Spots	The user selects a clickable surface area in an image (graph or table) GUI: Image with multiple clickable surface areas
OC	Open Constructive	The user provides 1 answer in the form of a text. This type of question is NOT COMPUTER SCORABLE. GUI: Free Text

The target for the item development teams was to develop an equal number of items for each content theme spread across the competency classes in the proportion 2:2:1. The distribution of items achieved is given in Table 3.

Table 3 Distribution of items between themes and classes

	Class 1	Class 2	Class 3
Quantity	**240**	**240**	**120**
Shape and space	**240**	**240**	**120**
Change and relationships	**240**	**240**	**120**
Uncertainty	**240**	**240**	**120**
Total	**960**	**960**	**480**

INULIS Assessment framework

The INULIS project adopted the PISA three-dimensional framework for assessment that distinguishes mathematical content, mathematical competence and situations. However, INULIS adjusted the relative importance it attaches to each dimension as shown in Table 4. Note this is different from the PISA weighting.

Table 4 INULIS A**ssessment Framework**

Dimension	**Weight**
Mathematical content	Major
Mathematical competence	Major
Situations	Major

Characterisation of mathematics in INULIS

The characterisation of mathematics using ‘overarching ideas’ is consistent with the INULIS project goals and this characterisation works well for the project. These themes are listed in Table 5.

Table 5 INULIS Themes

Mathematical Content
Quantity
Shape and space
Change and relationships
Uncertainty

Competency classes in INULIS

The use of mathematical competencies to organise the ‘quantitative knowledge and skills’ domain in the INULIS project is viable and appropriate. Here we apply the PISA competency classes as they were intended to the INULIS project domain. However, the weightings accorded competency classes in PISA i.e. 5:10:1 (2000) and 1:2:1 (2003) were revised for INULIS purposes. They were also adjusted for each type of test envisaged for INULIS i.e. tests for low skilled workers (Type 1) and junior executives (Type 2). Weightings for these tests have been adjusted to reflect INULIS priorities and they now stand at 8:2:0 and 2:6:2 respectively.

Table 6 INULIS Competencies classes and weightings

Competencies	Description	Type1 Test (weights)	Type2 Test (weights)
Class 1	Carrying out standard routine procedures	80%	20%
Class 2	Making connections and integrating ideas to assist problem solving	20%	60%
Class 3	Generalization, plan solution strategies and implement them in Problem settings, reflection		20%

Contexts and situations

INULIS follows the PISA approach quite closely as regards its interpretation of contexts and situations. The PISA approach in turn is very heavily influenced by the Realistic Mathematics Education (RME) movement which is based on problems set in real world contexts.

INULIS works with four broad contexts as identified in the PISA study viz. personal, occupational, public and scientific. In order to facilitate consistency in item production these contexts were further elaborated by the INULIS project team in Tables 7, 8, 9 and 10. Purely mathematical items are included in the scientific category where the context is mathematics itself. This intra-mathematical context is more evident in the INULIS item bank than in PISA items because INULIS attaches more weight to school mathematics in its development work.

Table 7 Outline of Personal Context

Recreation and Entertainment	Citizenship	Personal Finance	Retail
Recreation • Sports • Gym • Swimming • Travel Holiday Theatre/Cinema Eating Out /Bars Music /Entertainment	Personal Circumstances • identification • Interacting with Gov. Depts • Interaction with organisation on a personal level • Voting Critical citizenship • Interpreting • understanding	Household • maintenance • problem solving • planning • DIY • Casting • Wastage • Waste disposal Mortgages Car Loans Savings Pensions Healthcare - Medicines - insurance	Shopping Groceries Household Clothes Furniture Cars Houses Bills

Table 8 Outline of Public Context

Policies	Politics	Entitlements	Measurements/ Conversions
Education Medical Health Environment Demography Planning Local Planning National Transport Law and Regulations	Local and National Government	Tax Direct Tax Indirect Benefits: Medical card Child benefit Dole Grants	Units Measurements Conversions Road signage Tolls

Table 9 Outline of Occupational Context

The initial focus is on business but not exclusively (SME, Traders etc)
Business Farming Engineering Catering Building • Electric/Plumbing/Carpentry • Shapes/ Design • Quantity survey • Planning • Education • Retail Buying and selling

Table 10 Outline of Scientific Context

The scientific context includes anything that is scientifically or technologically oriented that involves mathematics.	
Glucose levels in blood	Calorific content of a diet
Heart rate and pulse	Pressure
Eyesight	Force
Emissions	Water flow
	Heat
Non-standard measures e.g. • population density • Man hours • Headage/unit **Note: I**ntra mathematics (Mathematics itself.) is also included	

INULIS Tests
All INULIS tests are generated by the user prior to testing using the INULIS system according to the directions issued by the INULIS project team. Only directions for INULIS Type 1and 2 tests are given. Tests that do not conform strictly to these directions are not recognised as INULIS tests.

Type 1 test specification and construction
The purpose of Type 1 tests is to measure that the level of quantitative knowledge and skills of low skilled workers in employment and those entering employment is adequate for employment. HR decisions based on test performance are likely to lead to further training.

Each Type 1 test has the following design features that reflect the INULIS rationale:

Mode: on-line
Duration: 60 minutes
Scoring: machine scored
Number of test items: 20
Mathematical themes: equal weight (5 items to each of 4 themes)
Mathematical competencies: reproduction (80%), connections (20%), reflection (0%)
Competency level: Reproduction: Lower (75%), Upper (25%)
Connections: Lower (50%), Upper (50%) in any order of themes
Real world contexts: balanced across the 4 categories but not in fixed proportions.

This test structure integrates the elements of the assessment framework already discussed in a way that can be operationalized by the INULIS partners.

The user can design a Type 1 test by using the INULIS system software to select 20 items from the database of test items. Each test item is selected using multiple identifiers to select an appropriate item. Individual items can be identified and selected for inclusion in a test by specifying a number of fields: name, theme, competency class, level, situation, and item-type. The user simply completes Table 11.

Table 11 INULIS Type 1 test structure

Competencies	Theme1	Theme2	Theme3	Theme4
Reproduction (80%)	**3 named items (Lower) 1 named items (Upper)**	**3 named items (Lower) 1 named items (Upper)**	**3 named items (Lower) 1 named items (Upper)**	**3 named items (Lower) 1 named items (Upper)**
Connections (20%)	**1 named item (Lower)**	**1 named item (Upper)**	**1 named item (Lower)**	**1 named item (Upper)**

Type 2 test specification and construction
The purpose of Type 2 tests is to measure that the level of quantitative knowledge and skills of workers in employment and those entering employment is adequate for entry at or promotion to junior executive level. HR decisions based on test performance are likely to lead to further training.

Each Type 2 test has the following design features that reflect the INULIS rationale:

Mode: on-line
Duration: 60 minutes
Scoring: machine scored mainly (with some input from HR user)
Number of test items: 20
Mathematical themes: equal weight (5 items to each of 4 themes)
Mathematical competencies: reproduction (20%), connections (60%), reflection (20%)
Competency level: Reproduction: Lower (50%), Upper (50%) in any order of themes
Connections: Lower (66.66%), Upper (33.33%)
Reflection: Lower (50%), Upper (50%) in any order of themes
Real world contexts: balanced across the 4 categories but not in fixed proportions.

This test structure integrates the elements of the assessment framework already discussed in a way that can be operationalized by the INULIS partners.

The user can design a Type 2 test by using the INULIS system software to select 20 items from the database of test items using the template provided in Table 12. Following the same procedures already used to construct a Type 1 test the template is completed. Each test item is selected using multiple identifiers to select an appropriate item. Individual items are identified and selected for inclusion in a test by specifying a number of fields: name, theme, competency class, level, situation, and item-type.

Table 12 INULIS Type 2 test structure

Competencies	Theme1	Theme2	Theme3	Theme4
Reproduction (20%)	1 named items (Lower)	1 named items (Upper)	1 named items (Lower)	1 named items (Upper)
Connections (60%)	2 named items (Lower) 1 named item (Upper)	2 named items (Lower) 1 named item (Upper)	2 named items (Lower) 1 named item (Upper)	2 named items (Lower) 1 named item (Upper)
Reflection (20%)	1 named item (Upper)	1 named item (Lower)	1 named item (Upper)	1 named item (Lower)

Test administration

Users are responsible for administering tests. When the user constructs a test for use then it may be used on-line in whatever configuration suits the user and testee, and circumstances permit e.g. for single employees/job applicants, small groups, or large groups. Users may wish to use the same test over a period of time or change the test for each group of testees.

The design of test items in the INULIS database and the test generation process combine to produce similar tests. INULIS makes no claims about exact comparability of tests constructed by users, however, great effort and care has been taken by the INULIS project team to ensure that all test items meet their internal specifications in terms of classification and grading and other project dimensions.

Test reports and interpretation

The system will generate an individual report for each testee when a test is taken. The report will be made available to the testee and the user. The user is responsible for interpreting test results for their own use and purposes. **INULIS accepts no responsibility for the interpretation of test results and subsequent uses by the user or the testee.** However, INULIS does offer some guidelines as to what it deems appropriate minimum standards in terms of test performance and quantitative knowledge and skills.

Type 1 test reports

Each individual Type 1 report specifies the testee's performance in two dimensions viz. number of correct responses by theme and competency class. The report also details the total number of correct responses in each theme and competency class**.** In addition a minimum overall standard is set as an overall percentage of correct responses/ total number of items. The system does not discriminate between non-attempts and incorrect responses. An illustration is given in Table 13.

Table 13 A sample Type 1 test report

Competencies	Level	Theme1	Theme2	Theme3	Theme4	Total	
Reproduction	Upper	A1		A3			
	Lower		A2		A4	A1+A2+A3+A4	
Connections	Upper	B1	B2	B3	B4	B1+B2+B3+B4	
	Lower	C1	C2	C3	C4	C1+C2+C3+C4	
	Total	A1+B1+C1	A2+B2+C2	A3+B3+C3	A4+B4+C4	(A+B+C)*	

Note 1: *A=A+A2+A3+A4 **Note 2: A1 and A3 must have value 0 or 1.**

***B=B1+B2+B3+B4**

***C=C1+C2+C3+C4**

INULIS Guideline: Testees scoring less than 50% correct items should be advised to repeat a similar test at another time after some further study and practice.

Type 2 test reports

Type 2 tests are used to assess suitability in terms of quantitative knowledge and skills of entrants/employees for supervisory/junior executive type duty. Each individual Type 2 report specifies the testee's performance in two dimensions viz. number of correct responses by theme and competency class. The report also details the total number of correct responses in each theme and competency class**.** In addition a minimum overall standard is set as an overall percentage of correct responses/ total number of items.

Unlike Type 1 tests, this test assesses all three competency levels (reproduction, connections, reflection).When the user interprets the results, it is important to realise that the competency classes form a hierarchy such that competencies at higher levels include the lower level competencies. Note that the system does not discriminate between non-attempts and incorrect responses. A template for a Type 2 test is given in Table 14.

Table 14 A sample Type 2 test report

Competencies	Level	Theme1	Theme2	Theme3	Theme4	Total
Reproduction	Upper	A1		A3		A1+A3
	Lower		A2		A4	A2+A4
						A
Connections	Upper	B1	B2	B3	B4	B
	Lower	C1	C2	C3	C4	C
						B+C
Reflection	Upper	D1		D3		
	Lower		D2		D4	D
	Total					(A+B+C+D)*

***A=A1+A2+A3+A4** ***D=D2+D4+D3+D4**

***B=B1+B2+B3+B4**

***C=C1+C2+C3+C4**

INULIS Guideline: Users may be reluctant to accept a test performance which does not achieve at least 50% correct responses.

Theoretical underpinnings

The INULIS project approach is informed by PISA which it acknowledges as a methodological guide. Since this is the case it is necessary and appropriate to offer a brief summary of the PISA framework as a backdrop to the INULIS project. A fuller discussion is available from the PISA website and from OECD/PISA publications (OECD, 2003; 2004).

The PISA framework

PISA attempts to test and compare mathematical knowledge and skills needed for life after school (not school mathematics) across 30 OECD countries and 11 partner countries. The focus on mathematical knowledge and skills for ‘life after school’ that inevitably includes ‘working life’ is especially relevant to the aims of INULIS. PISA (OECD, 2003:) defines a new construct called ‘mathematical literacy’ as:

… an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgements and to engage in mathematics, in ways that meet the needs of that individual's current and future life as a constructive, concerned and reflective citizen.

This capacity requires knowledge and application of traditional content of school mathematics and a range of process skills or competencies in mathematics including computation and mathematical thinking skills and strategies.

Mathematics content
The characterisation of mathematics that underpins the PISA project is heavily influenced by the Realistic Mathematics Movement (RME) (see deLange, 2001). Instead of looking at content in traditional topic areas, PISA examines content in terms of themes that might be experienced in everyday life e.g. quantity, space and shape, change and relationships, uncertainty. These themes are referred to as 'overarching ideas'.

Mathematical competencies
PISA recognises formal and technical mathematical skills but there is more. A set of eight skills was identified for PISA (2000/2003) including mathematical thinking, mathematical argumentation, modelling, representation, problem solving/posing, technical skills, use of aids and tools, communication. These skills clearly relate to the process of mathematics and are involved together in the performance of any mathematical task. Consequently, no attempt is made to assess these process skills or mathematical competencies as individual entities. These mathematical competencies are organised into a hierarchical structure of three competency clusters progressing from reproduction to connections and reflection. They are described as follows:

<u>Class 1:</u> Reproduction cluster; making simple computations and carrying out routine procedures
<u>Class 2:</u> Connection cluster; integration and connection of material from various overarching ideas to solve problems that are not simply routine but still involve quasi-familiar settings.
<u>Class 3:</u> Reflection cluster; mathematical thinking including generalization, insight, reflection.

Everyday situations
PISA works with four broad situations/contexts identified as personal, occupational, public and scientific. These are viewed as contexts in which mathematical problems arise in everyday life. A separate, fifth category, was used in the first cycle of PISA assessment (PISA 2000). This category was called 'intra-mathematical' in order to include purely mathematical items. This intra-mathematical context is now subsumed in the scientific category.

PISA assessment
PISA attempts to assess mathematical literacy using three dimensions viz. mathematical content, mathematical competencies, everyday situations or contexts.
Further, PISA identifies a number of aspects of mathematical literacy that it then uses to organise the domain for assessment purposes. Two major aspects are identified as *mathematical competencies* and *mathematical content*, while two minor aspects are called *mathematical curricular strands* and *situations and contexts*. In this scheme it is essential that a selection of overarching ideas (content mathematics) be made that reveals the essentials of mathematics in sufficient variety and depth.

A PISA test is constructed by selecting a number of overarching ideas, then devising a number of suitable test items in appropriate competency classes giving due attention to content coverage and everyday contexts.

Appendix A

Quality control issues

Each partner group engaged in item generation worked individually and collaboratively to produce test items using the assessment framework and structure, agreed item formats and publicly released PISA sample items for reference.

In addition, there were specially convened meetings (2) and other opportunities in INULIS project meetings (2) for the item production teams to share experiences, resolve differences and misunderstandings and generally to keep everyone working in unison with a shared understanding of the mission. Meanwhile the groups were in constant communication via email.

The item development teams followed a special regime to ensure quality and comparability of work between teams. A shared understanding of the assessment framework and its individual elements was developed in face-to-face meetings, and issues were teased out and committed to writing and later distributed to the teams for future reference. All items were developed in a common electronic format using agreed conventions.

Initially all items were produced in English and submitted in batches to the Dutch team for content clearance and then to the Irish team for English language clearance. All the Dutch items were cleared for content by the Irish team. Each team was responsible for implementing changes as required which could involve revisions or omission of submitted items. Omitted items were replaced and re-checked.

In this way all items were checked and re-checked for language, content, suitably etc. The items were exchanged electronically and in hard copy to facilitate the process. Finally, the project manager uploaded all amended and checked items electronically to the INULIS database. The respective teams produced Dutch and Greek language versions of the items in the database.

References

Cosgrave, J., Shiel, G., Sofraniou, N., Zastrutzki, S., and Shortt, F. (2004). *Education for life: The achievements of 15-year-olds in Ireland in the second cycle of PISA – summary report*. Dublin: Educational Research Centre.

De Lange, J. (2001) 'Mathematics for Literacy'. In B.L. Madison and L.A. Steen (Eds.), *Quantitative Literacy: Why Numeracy matters for Schools and Colleges*, 75-89. Princeton, NJ: National Council on Education and Disciplines.

De Lange, J. (1992) 'Assessing mathematical skills, understanding, and thinking.' In R. Lesh and S. J. Lamon (Eds.), *Assessment of Authentic Performance in School*

Mathematics, 195-214. Washington, DC: American Association for the Advancement of Science.
OECD (1999) *Measuring student knowledge and skills: A new framework for assessment*. Paris: Author.
OECD (2003) *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: Author.
Romberg, T. A. (2004) 'Classroom assessment studies'. In B. Clarke, D. M. Clarke, G. Emaluelsson, B. Johansson, D. V. Lambdkin, F. K. Lester, A. Wallby and K. Wallby (Eds.), *International Perspectives on Learning and Teaching Mathematics, 585- 600*. Goteborg, Sweden: National Centre for Mathematics Education.

Websites

A lot of documentation including technical reports are available at the OECD/PISA website. A small number is identified for special attention.

The PISA 2003 Assessment Framework
http://www.pisa.oecd.org/dataoecd/46/14/33694881.pdf

Measuring Student Knowledge and Skills
http://www.pisa.oecd.org/dataoecd/45/32/3369997.pdf

Sample Questions (Mathematics)
http://pisa-sq.acer.edu.au/showQuestio.php?testld=2296&questionld=1