



Progress Testing as Measurement of Learning

A psychometric approach

Jakob Wandall, University of Copenhagen & NordicMetrics.



Utrecht University



UNIVERSITY OF HELSINKI



KØBENHAVNS
UNIVERSITET



Norges miljø- og
biovitenskapelige
universitet



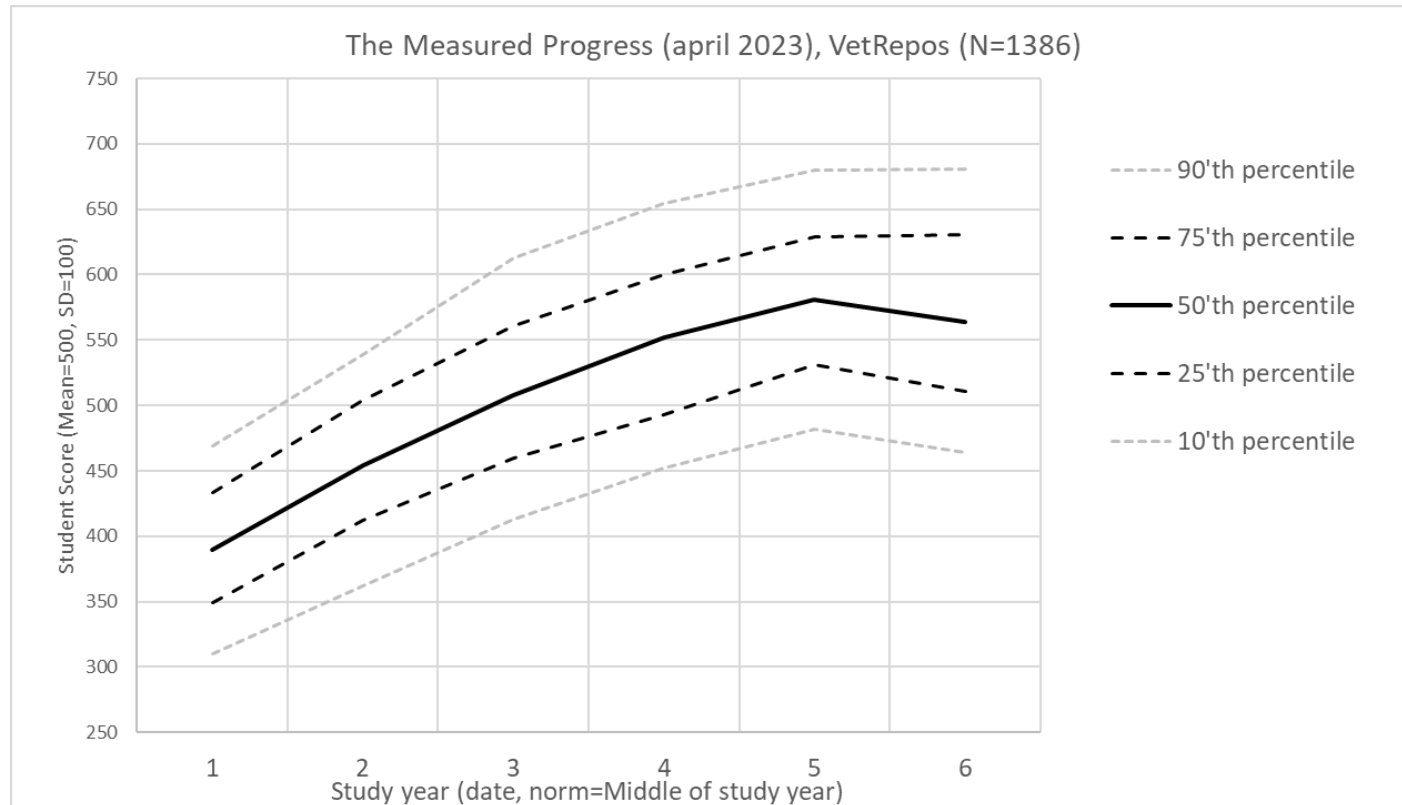


- **The project VetRepos 2021-2023 aims at:**

- Writing a lot of items that all is testing the same construct: Namely **Common veterinary skills and knowledge.**
- Establishment of a 2-level quality assurance system:
 1. QA of Content coverage (the Blueprint & QA-committee)
 2. Trailing all items – each 250-400 students – followed by statistical analysis of the responses: Does the items **measure the same construct & how difficult the single item are.**
- All approved test items are stored in an item bank with all information (content, difficulty etc.)
- The use of the test for measurement of progress: Formative testing, estimating student results and feedback to students and institution on the same scale across study years.



- **What do we actually mean by “learning”?**
- **Definition: The difference between two scores measured on the same scientific scale (interval level – we use the Rasch-Scale for analytic purposes)**
- For reporting we use the same scale as IEA (eg. TIMMS & Pirls) - a linear transformation of the Rasch scale: Mean = 500, Standard Deviation = 100.



Data behind the progress estimation

- Data from 4 item trials (N=1386)
- 150-300 responses pr study year

The students feedback:

- The scores (student performance)
- Their learning (student Progress)

What are measured (VetRepos Construct)

- Day 1 competences
- Not all knowledge and skills is tested - but everything is strongly correlated

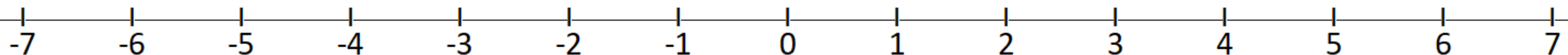
Interval scale - a requirement in *true* Progress Testing

- Most scales for the results in education are ordinal: Grading, ECTS, Percentiles etc. are all ordinal scales.
 - we **can** determine which student are performing the best in a specific test/exam.
 - we **cannot** compare results from different exams - not even in the same subject.
- Example: *In an anatomy test/exam - who is performing best? **The 2nd year student that gets a “B” OR The 5th year student that gets a “D”***
- In progress test: The same scale must be used for results **for all students in all study years**. The scale must be **unidimensional** and **equidistant**. In Ordinal scales **two scores on ordinal level cannot be subtracted**.
- Example: The ECTS scale.

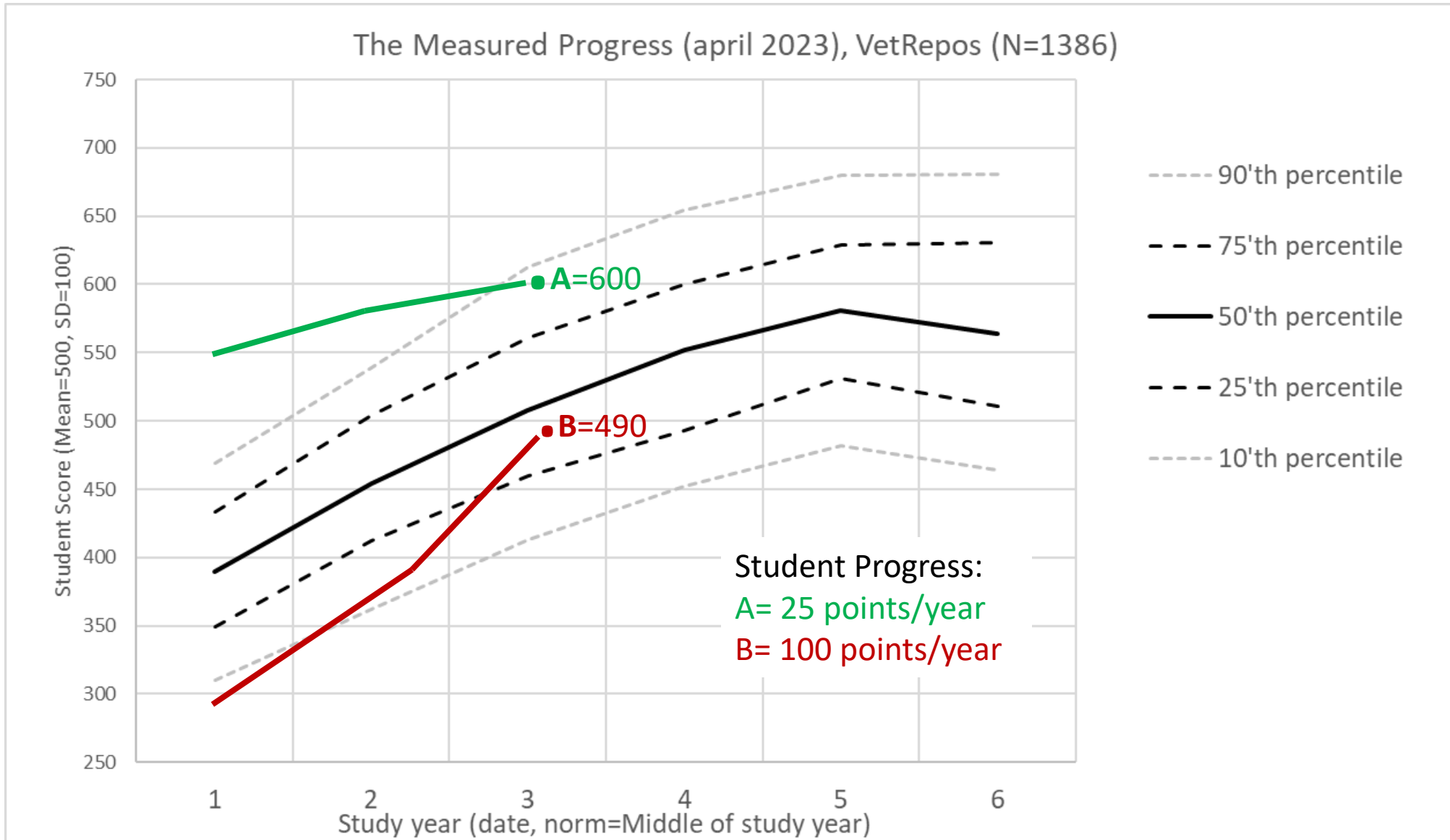
F	Fx	E	D	C	B	A
----------	-----------	----------	----------	----------	----------	----------

- Scientific measurement of progress e.g. the Rasch Scale. It goes from $-\infty$ to $+\infty$ (In practice, from -7 to +7) and “0” is located as the medium difficult item.

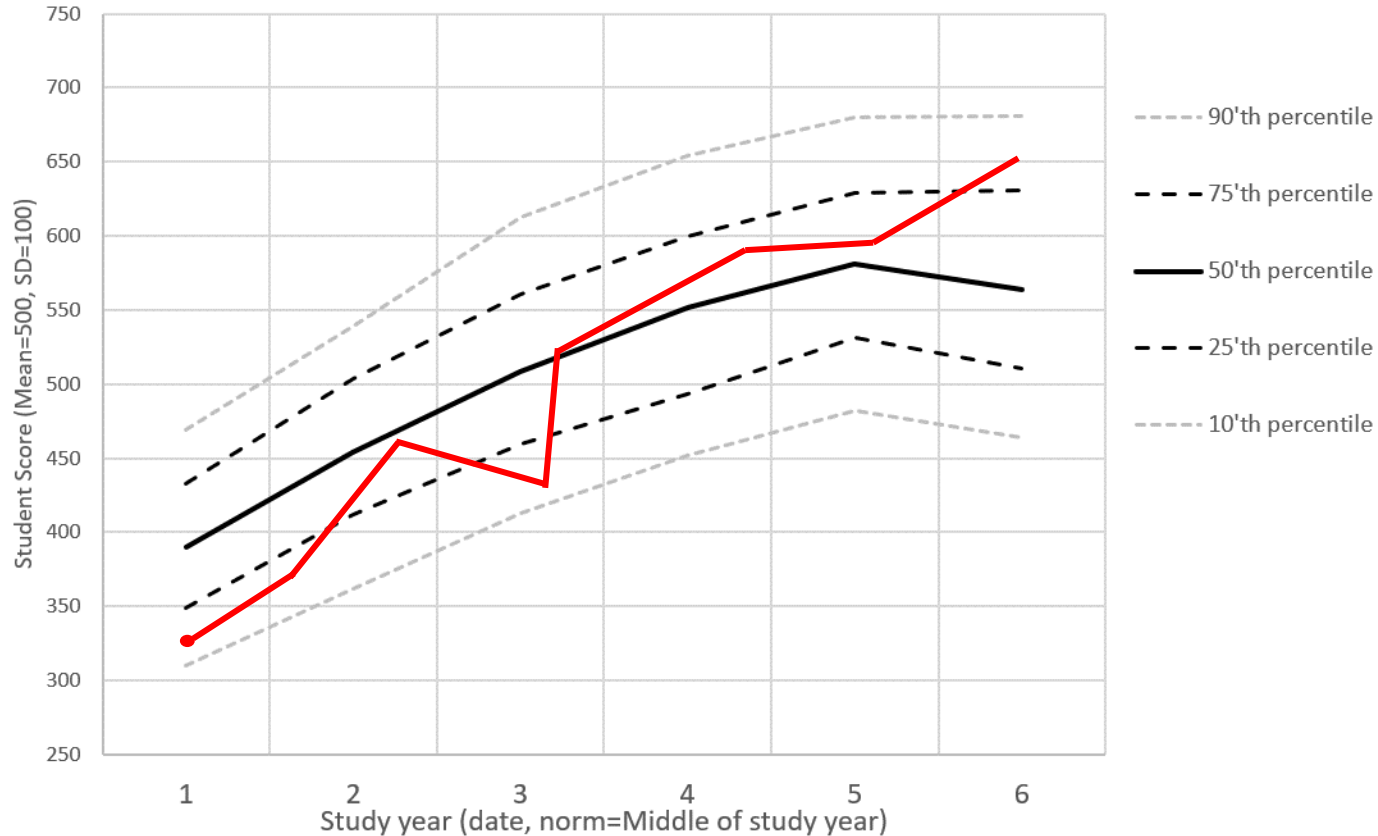
F	Fx	E	D	C	B	A
----------	-----------	----------	----------	----------	----------	----------



Who have learned the most? Student A or Student B



Student progress as a growth curve.



Similar curves can be made on each of the four subscales.

Student results as text and tables

Standardized text describing results compared to expectations (the norm in the performance data and the progress).

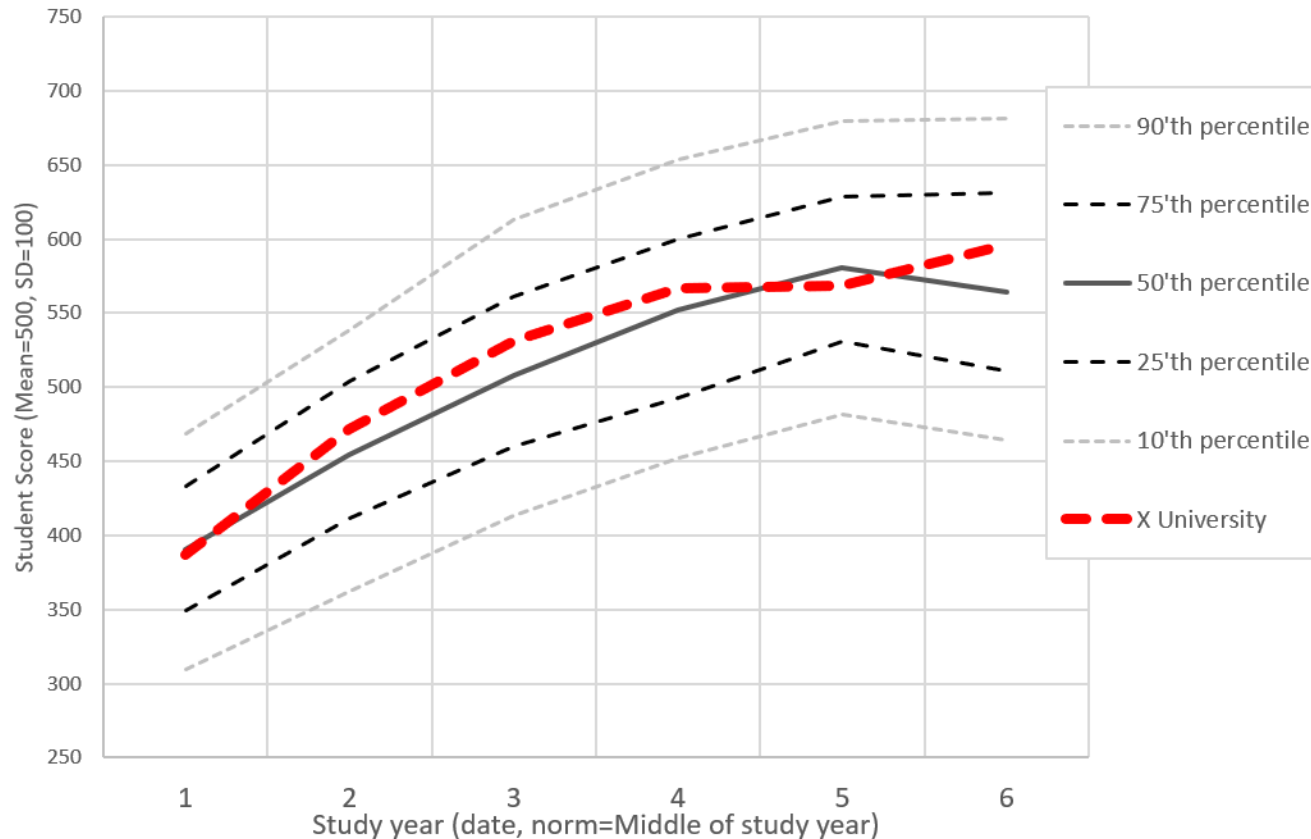
Tables with overview about performance and progress on more detailed results – e.g. scores on subscales/EAEVE areas.

Future possibilities

More **advanced feedback** based on the individual student's responses – e.g. based on AI- analysis of data.

Targeted intelligent **recommendations of study activities** based on the individual student's latest results and other student's experiences.

University average compared to the EU-average



Similar curves can be made on each of the four subscales.

Statistical Analysis (for internal use)

Tables with overview about average scores and progress on more detailed results – e.g. average scores on subscales/EAEVE areas.

More **advanced statistical analysis** based on comparison between the university score distribution and the total (EU) scores.

Caution: Handle test and data with care.

- **The VetRepos test is designed for formative purposes:**
 - Can be used 1-2 times pr. study year
 - The students like the test because it is low stakes and formative.
 - The student have no incentive to try to cheat or to inflate their test scores artificially in other ways – so the scores are probably highly reliable.
 - The test risk being compromised if the it is use in other ways fx for high stakes purposes like passing exams, qualification exams, admission tests or entrance exams.
 - Using the VetRepos for benchmarking of institutions can be done with caution, internally and low stakes purposes. If it is used for public league tables and ranking of institutions, it would probably distort the measures/scores.
 - The test is used for measurement of progress (formative testing, estimating student results and feedback to students and institutions).
 - Due to the lack of control/proctoring the test is fragile and sensible to misuse of test and testdata.