Why is standardisation important in assessments of SpLDs?

- SASC CONFERENCE JUNE 2023
- Rachel Simpson
- Chair of STEC

1

WHY <u>IS</u> STANDARDISATION IMPORTANT

2

UNIQUE FEATURES OF A DIAGNOSTIC ASSESSMENT

1. Looks for correlations between underdevelopment in specific skills and specific cognitive abilities in order to identify patterns of specific difficulty

2. Aligns those patterns of specific difficulty against recognised definitions of a specific learning difficulty

3. Applies a label which may have a lifelong impact upon the individual's self image and access to support

SASC
SpLD Assessment Standards Committee

3

What gives us confidence in our judgements?
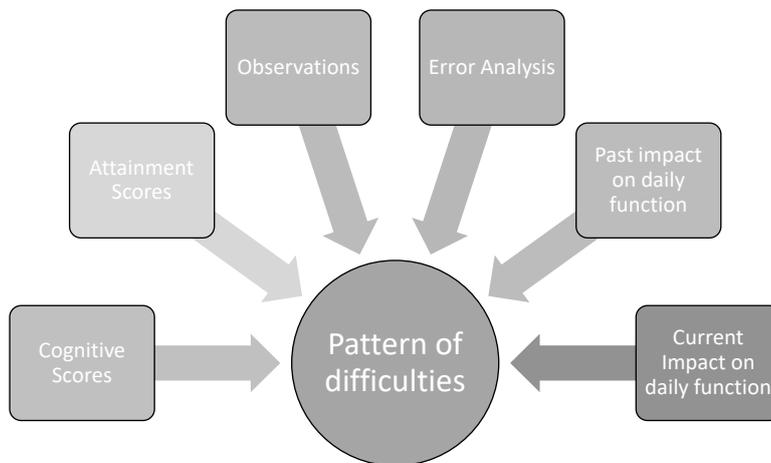
SASC
SpLD Assessment Standards Committee

4

1) Professional qualifications, knowledge and experience

- Recognised qualification and accreditation in diagnostic assessment
- Maintaining up-to-date CPD
- Reading all the latest guidance coming from our professional bodies
- Understanding SpLDs and using currently recognised definitions of SpLDs
- Being aware of a multiplicity of factors that can affect performance

SASC
SpLD Assessment Standards Committee

5

2) Holistic Data Gathering

SASC
SpLD Assessment Standards Committee

Observations

Error Analysis

Attainment Scores

Past impact on daily function

Cognitive Scores

Pattern of difficulties

Current Impact on daily function
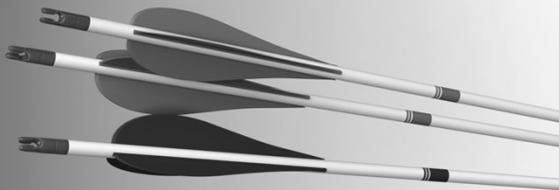
6

3) Using Psychometric Tests

The most effective way of quantifying performance is by using psychometric tests which

- Build theoretical models of those skills or abilities,
- Create tests to measure those models,
- Trial the tests on age cohorts
- Analyse the results
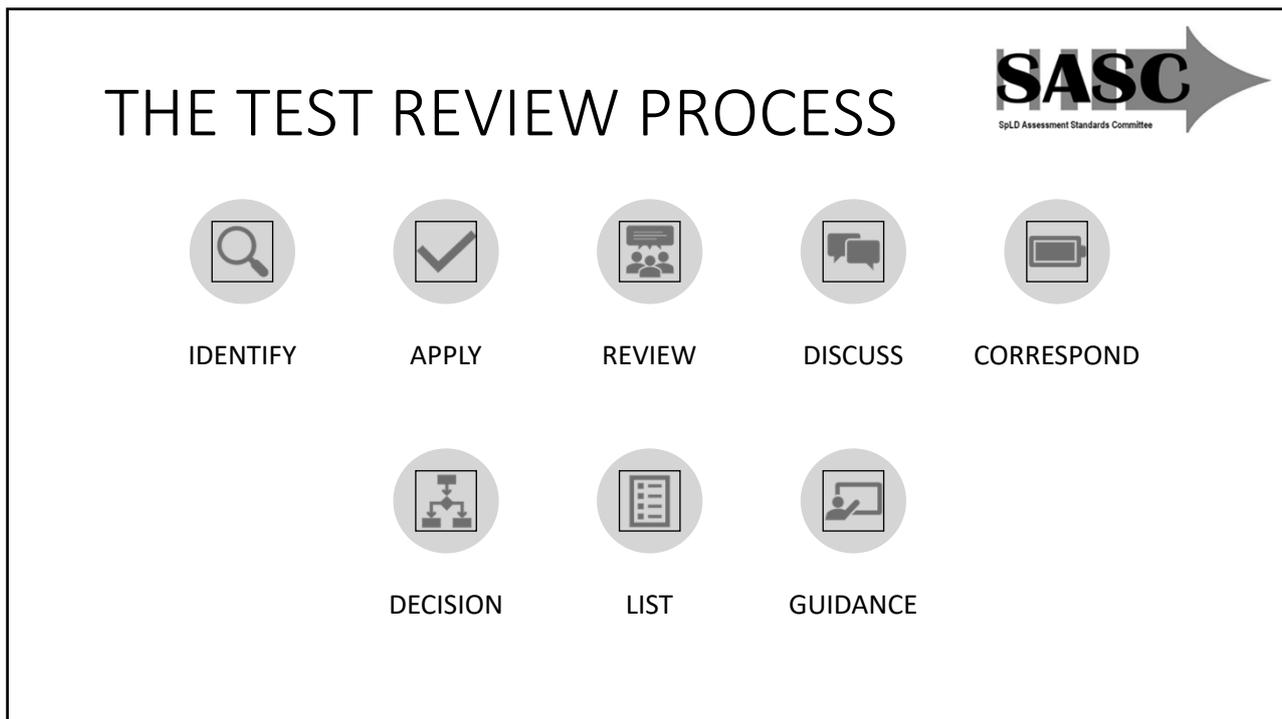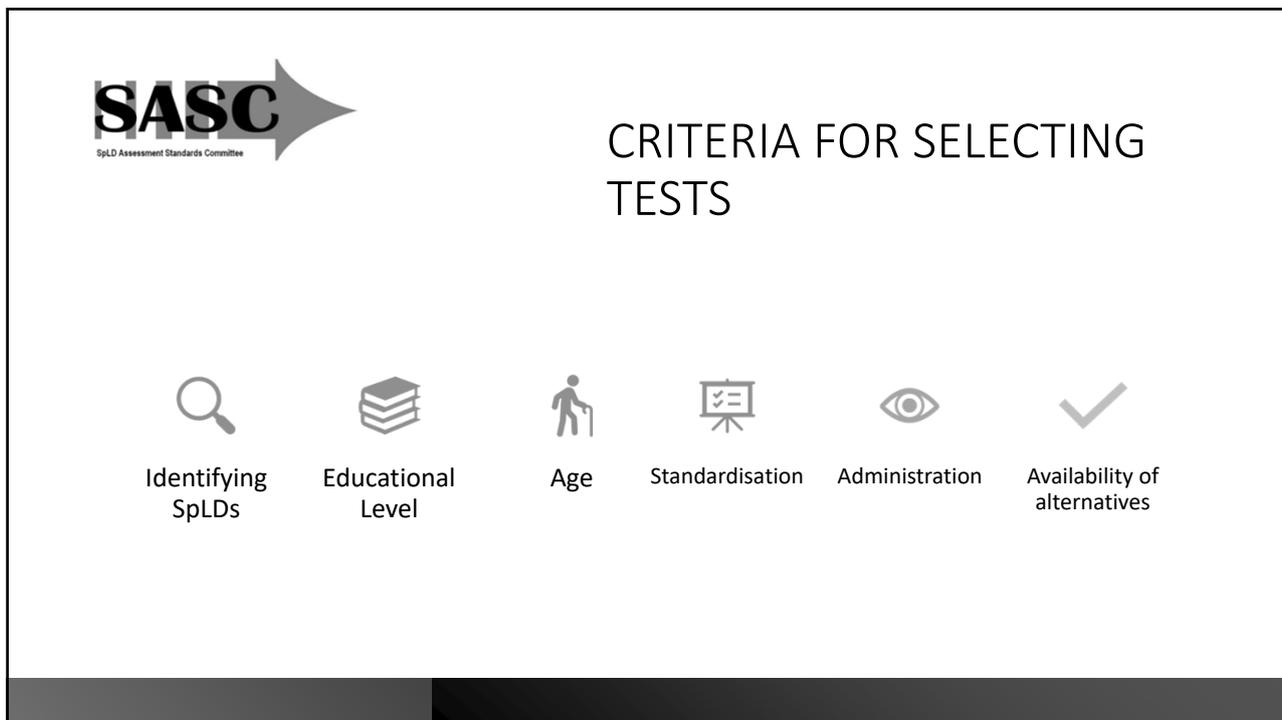- Translate raw scores into tables of norms

7

SASC
SpLD Assessment Standards Committee

HOW CAN WE ENSURE THAT THE TESTS WE USE DELIVER ACCURATE RESULTS?

8

THE TEST REVIEW PROCESS

IDENTIFY APPLY REVIEW DISCUSS CORRESPOND

DECISION LIST GUIDANCE

9

CRITERIA FOR SELECTING TESTS

Identifying SpLDs | Educational Level | Age | Standardisation | Administration | Availability of alternatives

10

IS IT APPROPRIATE FOR IDENTIFYING SpLDS?

What is its purpose?

- Screener?
- Dynamic assessment?
- Age/Grade Normed?
- Diagnosis of SpLD

What does it measure?

11

IS IT APPROPRIATE FOR THE EDUCATIONAL LEVEL?

What level is it designed for?

Does it accurately reflect required skills and abilities

12

AGE RANGE

What age range and why?

Does it accurately reflect required skills and abilities

13

EASE OF ADMINISTRATION

Clarity and Ease

Time

14

ARE OTHER TESTS AVAILABLE?

> Are other tests available that measure the same skills or abilities - sometimes necessary to include tests which may not be ideal but which fill, or partially fill, a gap in the range of available tests.
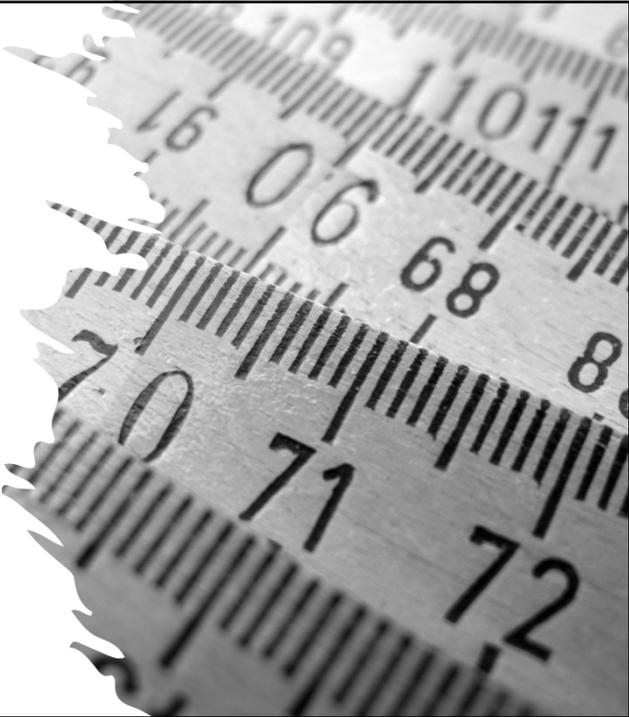
15

RIGOROUS STANDARDISATION

Theory

Sample

Validity & Reliability

Recency

16

WHAT IS
STANDARDISATION?

17

TYPES OF SCORE

- NORM REFERENCED
  - Compares an individual's skills or knowledge against the skills or knowledge of other individuals of the same age or grade. Scores may be Lickert scales, standard scores, scaled scores, percentiles, T scores or Stanines

- CRITERION REFERENCED
  - Compares an individual's performance against a pre-determined standard or goal (how much of that standard or goal has the individual mastered?). May not produce a score at all (may simply highlight which areas have not yet been mastered) or may produce a scale (numbers or grades) or a percentage score.

18

TYPICAL NORM REFERENCED SCORES

TYPE OF SCORE	MEAN	STANDARD DEVIATION	MIDDLE 2/3
Standard Score	100	± 15	85-115
Scaled Score	10	± 3	7-13
T Score	50	± 10	40-60
Stanine	5	± 2	3-7

19

SEMS AND CONFIDENCE INTERVALS

Why do we need to know them?

20

How are standardised tests developed?

Skills and abilities are intangible concepts which are not physically observable. They can only be measured through responses to items which require those skills or abilities to be applied to them.

Test design requires:

A theoretical model and rigorous evaluation of construct validity and content validity

Trial on a large sample

Re-testing on a smaller sample to establish test-retest reliability (would the test deliver the same results if administered at a different or by a different assessor)

Extensive, complex statistical calculations, to calculate norms and make those norms as accurate as possible

Comparison against other similar well-established tests to explore how well the results correlate

21

FACTOR ANALYSIS

• A theoretical construct is defined

• It is assumed that the construct will follow a normal distribution

• A set of items are designed to test that construct. These are sampled on a cohort.

• The results are analysed to looks for hidden patterns within a set of data.

• Those patterns may confirm the construct but may also identify certain items which ned to be removed because they disrupt the patterns.

22

ITEM RESPONSE THEORY

It is assumed that a trait follows a continuum which is not physically observable

The continuum will follow a normal distribution

A person's position on that continuum **can** be measured through their responses to a set of items which are organised to follow that continuum.

23

STRUCTURE EQUATION MODELLING

- A model is constructed as to how variables are causally related or associated.
- It is assumed that the variables will follow a normal distribution
- A test is developed to trial the model. It is expected that a pattern of connections will appear within the results
- A set of equations are created to work out how great or otherwise the expected causal relationships are – in order to check the model.

24

BOOTSTRAPPING

- A test is trialled on a sample
- A computer creates **multiple** new samples, all the same size, but each consisting of random combinations (including repeats or omissions) of the original data set.
- The mean, median and standard deviation are calculated for each data set.
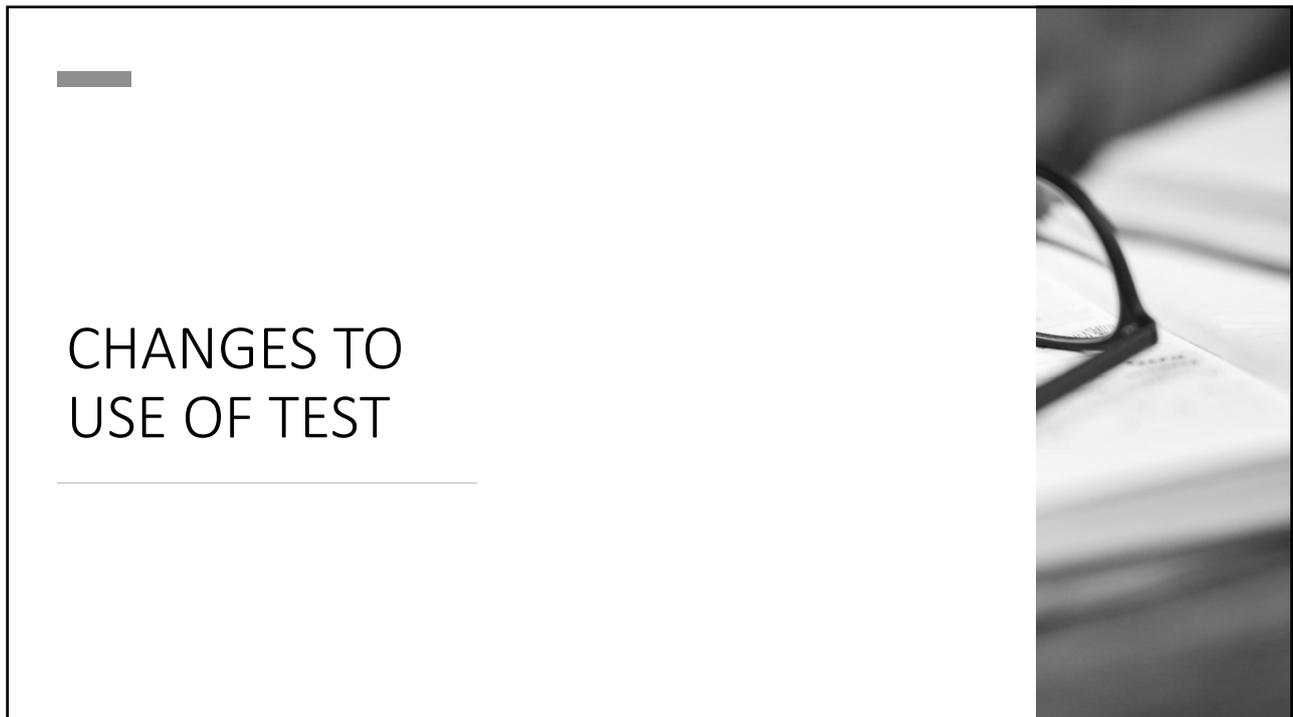- The resulting data distributions are used establish norms and confidence intervals

25

COMPUTERISED ADAPTIVE TESTING

A computer-administered test in which the algorithm selects the level of difficulty of successive items on the basis of the accuracy or otherwise of the individual's responses to the current item.

26

CHANGES TO
USE OF TEST

27

CONCLUSION

28