Seliant Mechanical Reasoning Test

Forms A and B | Six language versions

Technical Manual and User Guide

© 2026 Seliant ApS

Publisher Information

Seliant ApS
Gammel Hasserisvej 123
9000 Aalborg
Denmark

Copyright © 2026 Seliant ApS. All rights reserved.

Warning: No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the copyright owner.

Product Code: MECHREA26

Version: 1.0

For orders and inquiries:
Seliant ApS, Customer Service
Email: contact@seliant.com
Website: www.seliant.com

Seliant Mechanical Reasoning Test

Technical Manual

This document is the official Technical Manual for the Seliant Mechanical Reasoning Test. It gives a full overview of the purpose, structure and psychometric foundations of the assessment, with detailed guidance on administration, scoring and interpretation. The test measures the practical mechanical understanding needed in machine operation, assembly, maintenance and similar hands-on technical roles. That understanding of how gears, levers, pulleys, loads, liquids and simple circuits behave underpins safe and effective work with machines and equipment.

The test gives an objective and reliable measure of mechanical reasoning, supporting organizations in making informed, evidence-based decisions when selecting employees for mechanical trades. With two alternate forms and six language versions, it offers a standardized check of the basic mechanical understanding these roles rely on.

Besides setting out the theoretical basis and the evidence on reliability, validity and fairness, this manual gives practical guidance on administering the test, setting cut scores, meeting professional and legal standards, and interpreting results in line with best practice. Whether you are an HR professional, a recruiter, a hiring manager or another decision-maker involved in candidate evaluation, this manual is a key resource for understanding and applying the test.

© 2026 Seliant ApS

Contents

Test at a glance 1

1. Introduction and intended use 2

1.1 Purpose of the test 2

1.2 Intended audience 2

1.3 How to use this manual 2

2. What the test measures 3

2.1 Mechanical reasoning 3

2.2 Content areas 3

2.3 Question format 4

3. How the test was developed 5

4. Administration 6

5. Scoring and reporting 7

5.1 Raw score, percentile and band 7

5.2 Interpreting scores 8

5.3 The score report 9

6. Norms 10

7. Reliability and precision 12

8. Validity 13

9. Fairness 15

10. Using the test in selection 17

11. Test security and data quality 19

12. Language versions 21

Appendix A: Norm tables 22

Appendix B: Norm group composition 23

Appendix C: Decisions and adverse impact by cut score 24

Appendix D: Item statistics 25

Appendix E: Glossary 26

References 27

Version history 27

© 2026 Seliant ApS

Test at a glance

The Seliant Mechanical Reasoning Test is a 20-minute, 32-question test of practical mechanical understanding, built to screen applicants for mechanical trades.

Feature	Detail
Product code	MECHREA26
Manual version	1.0, November 2026
What it measures	Mechanical reasoning: working out how everyday mechanical and physical systems behave, from a picture
Intended use	Screening applicants for machine operator, assembly line, mechanic and similar hands-on technical roles
Not intended for	Engineering or other advanced technical roles, or separating the very highest scorers
Format	32 illustrated multiple-choice questions, each with two to four answer options
Forms	Two alternate forms, A and B, with no questions in common
Time	20 minutes; half of candidates finish within 12 minutes
Practice	Two untimed example questions before the test starts
Delivery	Online through Seliant's secure testing platform, in any modern browser, in full screen except on phones; nothing to install
Scoring	Number of correct answers, 0 to 32; no penalty for wrong answers
Reporting	Raw score, percentile and band (Lower, Middle or Upper), with a score chart and per-question results
Reliability	0.80 (Form A) and 0.81 (Form B); retest correlation 0.71
Norm group	5,460 results from about 5,000 people who took the test online, August to November 2026
Languages	English, Danish, German, French, Spanish and Portuguese, all scored on the English norms

© 2026 Seliant ApS

1. Introduction and intended use

The Seliant Mechanical Reasoning Test helps employers find applicants who understand how machines and physical systems behave: the practical sense that operating, assembling and maintaining equipment relies on.

1.1 Purpose of the test

Each question shows an everyday mechanical situation, such as gears, levers, pulleys, liquids, loads or a simple circuit. The candidate works out what will happen or which arrangement works best. This takes practical understanding, not formulas or calculation.

The test is designed to:

- Screen out, early in hiring, applicants who clearly lack the basic mechanical reasoning a job needs, before costlier steps such as interviews or practical trade tests.
- Compare candidates on an equal footing, with the same instructions, time limit and scoring for everyone and a norm group of 5,460 results.
- Add an objective measure to other information such as an interview, work history or a practical trade test.

A test score should not be the only basis for a hiring decision (Chapter 10).

1.2 Intended audience

The test is for applicants to hands-on technical roles that involve operating, assembling, maintaining or repairing machines and equipment, such as:

- Machine and production operators
- Assembly line workers
- Mechanics
- Maintenance technicians
- Apprentices and trainees entering these trades

The test is pitched at the level these roles need. In the norm group, the main trade groups average 21.0 to 21.8 out of 32, close to the overall average of 21.2 (Chapter 6).

Who it is not for. Engineers and other advanced technical roles. The test does not separate the very highest scorers from each other; it measures most precisely around the scores where screening decisions are made.

1.3 How to use this manual

Chapters 2 and 3 describe what the test measures and how it was built. Chapters 4 and 5 cover administration, scoring and how to interpret scores. Chapters 6 to 9 give the norms and the evidence on reliability, validity and fairness. Chapter 10 gives guidance on using scores in selection, Chapter 11 covers security and Chapter 12 the language versions. The appendices hold the full tables.

© 2026 Seliant ApS

2. What the test measures

The test measures mechanical reasoning: the ability to work out, from a picture, how a physical or mechanical system will behave.

2.1 Mechanical reasoning

Candidates see a situation and predict its outcome: which way a gear turns, where a beam balances, which pulley arrangement needs the least effort, which container empties fastest. The questions draw on an intuitive grasp of physical principles, of the kind built up through hands-on work and everyday experience. They do not require calculation or formal physics training, and the test is not a speed test.

2.2 Content areas

The questions cover 13 areas of practical mechanics, with forces and motion the largest on both forms. The forms differ most in heat (5 questions on Form A, 2 on Form B) and in liquids and pressure (3 and 5), but they measure the same ability at the same difficulty (Chapter 8).

Content area	What the questions cover	Form A	Form B
Forces and motion	How things start, stop, fall, swing, turn and fly through the air, and the forces on beams, vehicles, boats and aircraft	8	8
Liquids and pressure	Pressure at depth, liquid levels in connected containers, flow through pipes, hydraulic lifts and air pressure	3	5
Heat	How heat moves through materials, why some feel colder than others, expansion when heated, and changes such as freezing	5	2
Gears and rotation	The direction and relative speed of connected gears, and how rotation passes from one gear to the next	3	3
Levers and balance	Which side of a lever or beam goes down, where to place a support, and how distance from the pivot changes the effort	3	3
Pulleys and mechanical advantage	How pulley systems share a load, the force needed to lift it and how far the rope must be pulled	2	2
Tools and everyday objects	How handle length changes turning force, how grip depends on friction, and easier ways to lift and carry loads	2	2
Center of gravity	Where an object's center of gravity lies and whether it stays upright, from stacked objects to boats	1	2
Light	How light travels: shadows, reflections in mirrors, and light compared with sound	2	1
Electricity	Simple circuits with batteries and bulbs, and what happens when part of a circuit changes	1	1
Magnets	How magnets attract, repel and line up, and which materials a magnetic field passes through	1	1

© 2026 Seliant ApS

Content area	What the questions cover	Form A	Form B
Springs	How springs stretch under a load and share it between them	1	1
Spatial visualization	Following in the mind's eye how movement passes through linked parts	0	1
Total		32	32

The descriptions show the kinds of situations covered; the questions themselves stay confidential (Chapter 11).

2.3 Question format

Each question has a picture, a short question in plain workshop language, and two to four answer options. Many have only two or three options because the physics allows only that many outcomes: a gear turns one way or the other, a beam tips left or right, or nothing changes. Form A has 1 two-option, 17 three-option and 14 four-option questions. Form B has 14 three-option and 18 four-option questions.

What this means for scores. Because of the two- and three-option questions, someone answering at random would average about 9 or 10 correct, not 8. A score below 16 cannot be reliably told apart from guessing, which is why 16 is the lowest recommended cut score (Chapter 10).

© 2026 Seliant ApS

3. How the test was developed

The two published forms were built from a 66-question bank calibrated on 2,894 candidates, revised, and then confirmed on a further 5,460 test results.

1. **Writing the questions.** Questions were written to cover the 13 content areas, each with a clean vector illustration and short, plain wording. Every keyed answer was checked against the physics of the situation and against the drawing itself, including measuring the illustration where the answer depends on proportions, so that only one answer is defensible.
2. **First calibration.** The bank was given online as two 33-question forms to 2,894 candidates. Each question was analyzed for difficulty, discrimination (how well it separates stronger from weaker candidates), how its wrong answers behaved, and fit to an item response model. On every question, the keyed answer attracted the strongest candidates, confirming the keys.
3. **Revision.** One question was removed from each form. One was too hard and drew most candidates to a single wrong answer; the other separated stronger from weaker candidates poorly. Three pairs of questions were swapped between the forms to make them equally difficult, since the original forms had differed by 1.4 raw points. Seven questions that discriminated less well were reviewed and, where needed, reworded or redrawn.
4. **Confirmation.** The revised 32-question forms were taken 5,460 times, by about 5,000 people, between 20 August and 3 November 2026. This sample supplies all the statistics and norms in this manual. The two forms now differ by 0.04 of a raw point on average.
5. **Final corrections.** On 4 October 2026, the illustration of one question was clarified and the wording of another was aligned with its picture. Both became slightly easier, raising average scores by about 0.05 of a point, too little to affect the norms; no other question changed.

In the platform, Forms A and B may appear as A-v2 and B-v2.

© 2026 Seliant ApS

4. Administration

The test is taken online in one sitting: instructions and two practice questions, then 32 questions with a 20-minute limit.

Delivery. Candidates take the test online through Seliant's secure testing platform, in English, Danish, German, French, Spanish or Portuguese; they should use the language they read best (Chapter 12). Nothing needs installing: the test runs in any modern browser, such as Chrome, Edge, Firefox or Safari, with JavaScript and cookies enabled and a stable broadband connection. Pop-up blockers should be turned off if they interfere.

Device. A computer or tablet is recommended, ideally with a screen of 13 inches or more. Candidates on phones scored 1.5 points lower on average; the gap was spread evenly across the questions and did not depend on screen size, so its cause is unclear, but a larger screen is the safer choice.

What candidates see.

1. Instructions: the number of questions, the time limit and how to answer.
2. A practice test: two example questions, untimed, where candidates can check their answer and read an explanation.
3. The test: 32 questions shown one at a time, with a timer on screen. Candidates can move back and forth between questions. On the last question, End finishes the test; when the 20 minutes are up, the test ends automatically.

Before and during the test. Candidates confirm their identity before they start; except on phones, the test then switches to full screen, which they must stay in to complete it. They are told that wrong answers are not penalized, so it is better to answer every question than to leave one blank. They should sit somewhere quiet without interruptions, close other applications and browser tabs, and work without internet searches or help from others; the platform records signs of outside help, such as leaving the test window (Chapter 11).

Timing. The limit is generous: half of candidates finish within 12 minutes, 90% within 17, and 98% before the limit. Most of those who reach the limit have still answered every question.

Accommodations. Reasonable accommodations can be arranged for candidates with documented needs. Extra time can be given where a disability affects a candidate's pace; since almost everyone finishes in time, it has little effect on comparability. The questions depend on diagrams, so the test is not suitable for candidates who cannot see them, including screen-reader users; use another assessment. No question depends on telling colors apart, so color blindness is not a barrier.

Retesting. When an employer initiates a retest, the platform automatically gives the candidate the other form, and no minimum waiting period applies. A retest on the other form raises scores by 0.6 points on average, while retaking the same form raises them by about 3 points because the questions are familiar (Chapter 11).

© 2026 Seliant ApS

5. Scoring and reporting

A candidate's score is the number of questions answered correctly, from 0 to 32, reported as a raw score, a percentile and a band.

5.1 Raw score, percentile and band

Raw score. One point per correct answer, with no penalty for wrong answers. An unanswered question counts as wrong; 2% of candidates leave a question unanswered.

The two forms. Forms A and B have the same average score (21.2) and the same reliability, so a raw score means nearly the same on either form. Form B's scores are slightly more spread out, so percentiles are read from each form's own column in Appendix A. The two columns differ by at most 3 percentile points.

Percentile. The percentage of the norm group who scored lower, counting half of those with the same score. A percentile of 60 means the candidate did better than about 60% of the norm group.

Bands. Scores are grouped into three bands:

Band	Raw score	Share of norm group
Lower	0–17	25%
Middle	18–25	51%
Upper	26–32	24%

Bands describe standing, not pass or fail: the Lower band includes scores of 16 and 17, which pass the lowest recommended cut score.

© 2026 Seliant ApS

5.2 Interpreting scores

Read a score against the norm group named on the report: a percentile shows how the candidate compares with that group, not with your own applicants (Chapter 6).

Raw score	Band	Percentile, Form A	Percentile, Form B	What it means
0–15	Lower	Under 1 to 13	Under 1 to 15	Below the lowest recommended cut. A score this low cannot be reliably told apart from guessing, so it says little about the candidate's mechanical reasoning. The lowest 16% of the norm group score here.
16–17	Lower	17–21	19–24	Above guessing level, but a weaker practical grasp of how machines and physical systems behave than most of the norm group. 9% of the norm group score here.
18–25	Middle	26–74	29–71	A solid practical grasp of how machines and physical systems behave. Half of the norm group score in this band.
26–32	Upper	80 to over 99	78 to over 99	A strong practical grasp, ahead of at least three in four of the norm group. The test separates the very highest scores less finely (Chapter 7).

What candidates get right. The share of easy, medium and hard questions answered correctly rises steadily with the score (both forms combined). Easy questions are those that 80% or more of the norm group answer correctly, and hard ones those that fewer than half do (Appendix D).

Raw score	Easy questions	Medium questions	Hard questions
0–15	62%	36%	17%
16–17	76%	49%	21%
18–25	88%	67%	34%
26–32	97%	88%	66%

Examples.

- A score of 22 on Form A is at the 54th percentile, in the Middle band: the candidate did better than about half of the norm group.
- A score of 27 on Form B is at the 84th percentile, in the Upper band: better than about five in six of the norm group.
- A score of 17 on Form A is at the 21st percentile, in the Lower band. It passes the lowest recommended cut of 16, but the candidate did better than only about one in five of the norm group.
- A score of 15 on either form is below the lowest recommended cut and cannot be reliably told apart from guessing.

How precise a score is. Every test score carries measurement error. Here the standard error is about 2.3 to 2.4 raw points, so a score could easily come out 2 points higher or lower on

© 2026 Seliant ApS

another occasion. In percentile terms the uncertainty is larger: between two sittings, a candidate's percentile moves by a median of 12 points. On a retest with the other form, 68% of candidates stay in the same band and 99% within one band, but only 10% get exactly the same raw score.

Comparing candidates. Treat differences of fewer than 5 raw points as ties. With a 5-point gap there is a 92% chance the higher scorer is genuinely stronger; with a 2-point gap, only 71%.

5.3 The score report

The employer's report for each candidate has two pages:

- **Overall score:** the raw score, the percentile against the norm group named on the report and the band, with a short explanation.
- **Score distribution:** a chart of how many people in the norm group got each raw score, with the candidate's score marked and the three bands shaded.
- **Test details:** form, language, items answered and time used.
- **Per-question results:** each question's content area and whether it was answered correctly, wrongly or not at all.

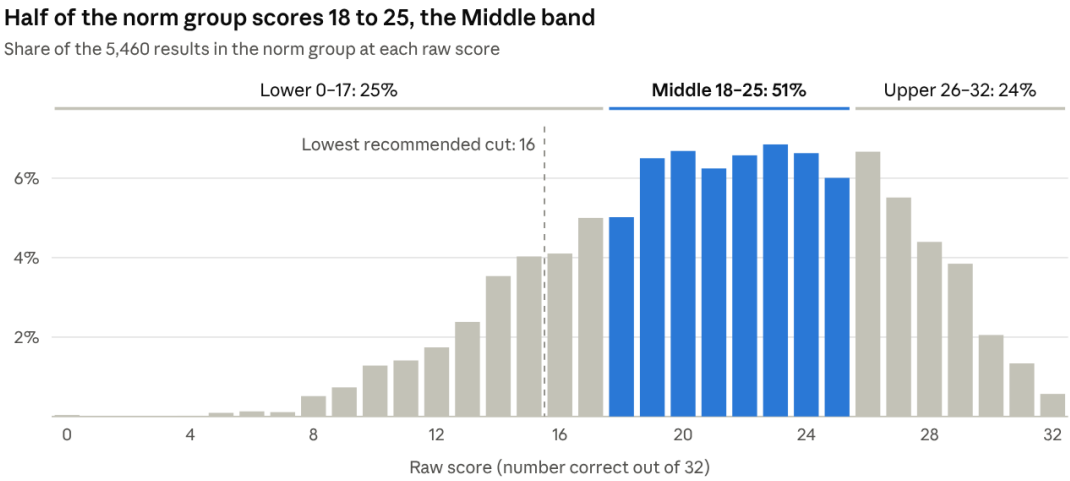
Integrity flags are not printed on the report; employers see them in the verification note on the candidate's page in the platform (Chapter 11).

Reading the per-question results. Each content area has only 1 to 8 questions on a form, so a wrong answer or two says little about a candidate's grasp of that topic. Base decisions on the total score, not on single questions or areas.

© 2026 Seliant ApS

6. Norms

Scores are compared with a general norm group: 5,460 results from about 5,000 people who took the test online between August and November 2026.



Norm group of 5,460, both forms combined · August to November 2026

Most scores fall between 18 and 27. The lowest recommended cut, 16, falls inside the Lower band and screens out about 16% of the group.

Who is in it. People who chose to take the test online, each counted once per form; they are not job applicants, and they took the test before full screen was required (Chapter 4). 71% are men and 19% women, and the median age is 26. 48% live in the US, and about half of the rest in Australia, Canada, India and the UK. Half are employed, and about half give their job field as engineering, manufacturing, construction or automotive. Appendix B gives the full composition.

What this means. An employer's applicants may score higher or lower than this online group, so a percentile describes standing in a broad online sample, not among applicants for a particular job. Band boundaries and cut scores are set in raw-score terms, so they stay the same whoever applies.

Norm tables. Appendix A gives the general percentile for each raw score on each form. The same tables apply to all six language versions (Chapter 12).

Occupational groups. The main trade groups score much like the general group: their averages run from 21.0 (manufacturing) to 21.8 (construction), against 21.2 overall. Among job titles, machine operators averaged 19.9 (183 people) and automotive mechanics and technicians 21.8 (262), so the general norm can place a machine operator applicant somewhat lower than their own peers would. Occupational norms are not published in this version.

Selecting a norm group. Where the platform offers more than one norm group for this test, select one for each job and compare every applicant to that job with it; this manual documents the general norm group. Bands and cut scores are set in raw scores, so they stay the same whichever norm group is used.

© 2026 Seliant ApS

Local norms. An employer that tests many applicants can build its own norms once it has results for at least 300 applicants to similar roles; that is roughly the number needed to pin a percentile to within 5 points.

© 2026 Seliant ApS

7. Reliability and precision

Scores are consistent enough for screening: internal consistency is 0.80 on Form A and 0.81 on Form B, and 86% of retested candidates got the same pass/fail verdict at a cut score of 16.

Measure	Form A	Form B
Candidates	2,637	2,823
Average score (SD)	21.2 (5.2)	21.2 (5.5)
Internal consistency (KR-20)	0.80	0.81
95% confidence interval	0.79–0.81	0.80–0.82
Split-half reliability	0.80	0.81
Standard error of measurement	2.3 points	2.4 points

Internal consistency shows how consistently the 32 questions measure the same ability. The two forms are equally reliable; the difference between them is not significant.

Retests. 270 candidates took both forms. Their two scores correlated 0.71, or 0.73 after allowing for retakers' narrower spread of scores. Their second score was 0.6 points higher on average, a small practice effect. This, rather than internal consistency, is the better guide to how much a score can move between sittings. Retaking the same form is different: the 68 candidates who did so gained 3.2 points on average, and their two scores correlated only 0.56. Retests should therefore always use the other form.

Where the test is most precise. Precision is highest around raw scores of 14 to 17, near the recommended cut scores, and lower at the top of the range. This is by design: the test does not need to separate the very highest scorers.

Pass/fail decisions. A cut score of 16 screens out the lowest 16% of the norm group. At that cut, 86% of real retests and 89% of simulated retests give the same verdict. At a cut of 18, which screens out 25%, the figures are 81% and 86%. Appendix C gives these figures for every cut from 14 to 21, with the estimated share of candidates wrongly rejected or wrongly accepted.

© 2026 Seliant ApS

8. Validity

The test measures one coherent ability, and its relations with other variables are consistent with mechanical reasoning. Evidence that scores predict job performance comes from published studies of similar tests, and employers should confirm it in their own setting.

Content. The questions sample 13 areas of practical mechanics that operating, assembling and maintaining equipment draws on (Chapter 2). They are written in plain language, with every key checked against both the physics and the drawing (Chapter 3).

Internal structure.

- One ability dominates: the largest factor in the questions' correlations is 3.4 to 3.5 times the size of the next.
- Every question fits an item response model closely: on average, observed results differ from the model's predictions by 6 percentage points or less.
- On every question, the keyed answer attracts the strongest candidates.
- Forms A and B measure the same ability at the same difficulty (Chapter 5).

Relations with other variables. In the norm group:

- Scores rise slightly with age (correlation 0.12) and with education: candidates with a four-year degree average 22.2, high school graduates 20.9.
- Occupational groups differ by up to 2.5 points, from 19.5 (health care) to 22.0 (information technology), with the trades in the middle (Chapter 6).
- Men outscore women by 0.62 standard deviations. That is somewhat smaller than the 0.76 reported for the mechanical reasoning test of the Differential Aptitude Tests in large samples (Hyde, 2005); see Chapter 9.

Job performance. No study yet links scores on this test to job performance or training results. Published research on tests of the same kind shows they predict both:

- In manufacturing jobs making electro-mechanical components, a mechanical comprehension test was the best single predictor of supervisors' performance ratings (r = .38 for 193 employees; Muchinsky, 1993).
- In a meta-analysis of firefighter selection, mechanical comprehension tests predicted job performance (r = .26 across 3,087 people; .54 after correcting for unreliable ratings and restricted range) and training results (r = .37; .62 corrected) (Barrett, Polomsky and McDaniel, 1999).

These results support using a mechanical reasoning test for mechanical work. They come from other tests and jobs, though, so they do not show how well this test predicts performance in a particular employer's roles.

Local validation. Employers using the test for pass/fail decisions, especially in the US, should check how scores relate to performance in their own jobs. The quickest route is to test current employees in the role and compare their scores with supervisors' performance ratings; 100 or

© 2026 Seliant ApS

more people give a usable answer. Comparing hires' scores with later performance also works, but it takes longer and understates the relationship, because people below the cut were never hired.

© 2026 Seliant ApS

9. Fairness

Most questions work the same way across groups, but average scores differ by sex and by ethnicity, so any cut score will pass some groups at lower rates than others.

Group averages. From the 5,272 results whose background answers were confirmed as truthful; ethnicity for US residents only, all self-reported.

Group	Candidates	Average score	Difference (SD units)
Men	3,775	21.8	reference
Women	1,005	18.6	−0.62
White (US)	1,267	22.3	reference
Black or African American (US)	280	17.5	−0.96
Hispanic or Latino (US)	413	20.3	−0.40
Asian (US)	86	21.1	−0.24
Native American or Alaska Native (US)	50	19.3	−0.62

Sex. Men average 3.2 points more than women. Mechanical reasoning tests have long shown differences of this size (Chapter 8); here it appears on both forms and within the trade occupations.

Ethnicity. Among US candidates, Black candidates average 4.8 points below White candidates and Hispanic candidates about 2 points below. The gaps remain after allowing for sex, age, education and device. A Black–White difference of about one standard deviation is in line with large studies of cognitive ability tests used in employment (Roth et al., 2001).

Question-level checks. Every question was checked for bias by comparing groups with the same overall score. Most show no meaningful difference: 54 of 64 for women and men, 59 for Black and White candidates, and 58 for Hispanic and White candidates. The rest show moderate or large differences, in both directions. Removing the three questions with the largest difference against women would reduce the sex gap only from 0.62 to 0.58, so the sex gap is spread across the test rather than caused by a few questions.

Adverse impact. At a cut score of 16, women pass at 0.80 of the men's rate, right on the four-fifths line; allowing for sampling error (0.77–0.84), it could fall on either side of the line among an employer's own applicants. Black candidates pass at 0.68 of the White rate and Hispanic candidates at 0.91. Native American and Alaska Native candidates, a small group of 50, pass at about 0.82 (95% interval 0.68–0.95). Higher cuts widen these gaps; Appendix C gives the ratios for each cut, and Chapter 10 explains what they mean for setting one.

Language. Candidates whose first language is not English score lower on the English version, and more so the weaker their English. Compared with native speakers of the same sex, age, education and device, the gap is 0.6 points for those who say they are fluent, 1.7 for "very well" and 2.4 for "not very well". A few questions depend more on reading English than others. Candidates who do not read English fluently should take the test in their own language where it is one of the six versions (Chapter 12).

© 2026 Seliant ApS

Device. Candidates on phones scored 1.5 points lower than those on computers. The gap was spread evenly across the questions and unrelated to screen size; a computer or tablet is recommended.

Age. Scores rise slightly with age (correlation 0.12).

© 2026 Seliant ApS

10. Using the test in selection

Use the test to screen out applicants who clearly lack basic mechanical reasoning, not to rank close candidates. Set any cut score at 16 or above, base it on what the job needs, and check its adverse impact.

As one input. Use the score together with other evidence, such as an interview, work history, a practical trade test and references. One approach is to use the test as an early screen and choose among those who pass on other grounds.

Setting a cut score.

1. **Start from the job.** With people who know the work, decide what level of mechanical understanding a new hire needs. Express it as a raw score rather than a percentile: raw scores stay fixed, while your applicants may differ from the norm group.
2. **Go no lower than 16.** Below 16, a candidate cannot be reliably told apart from someone guessing. About 1 in 100 pure guessers reach 16, and 2 to 3 in 100 reach 15.
3. **Weigh the job's needs against adverse impact.** Higher cuts screen out more people, misclassify slightly more of them, and in the norm group widen the pass-rate gaps between groups. At 16 the women-to-men ratio is 0.80 and the Black–White ratio 0.68; at 18 they are 0.73 and 0.57 (Appendix C).
4. **Review it in use.** Track pass rates by group among your own applicants, and revisit the cut if they differ from expectations.

Comparing candidates. Treat scores fewer than 5 points apart as equivalent, and use other information to choose between them (Chapter 5).

The US legal framework. This is general information, not legal advice; consult employment counsel before setting a cut score.

- Under the Uniform Guidelines on Employee Selection Procedures, a selection rate below four-fifths (80%) of the highest group's rate is generally treated as evidence of adverse impact, and smaller differences can also count when they are significant in both statistical and practical terms (29 CFR 1607.4).
- A selection procedure with adverse impact is treated as discriminatory unless the employer can show that it is valid for the job, normally through a validation study of this test in that job (29 CFR 1607.3). The published studies in Chapter 8 concern other tests, so on their own they do not show this. Employers should also look for equally valid alternatives with less adverse impact.
- Title VII, as amended by the Civil Rights Act of 1991, bars adjusting scores, using different cut scores, or otherwise altering test results by race, color, religion, sex or national origin, so every applicant for a job is compared with the same norms and the same cut score (EEOC guidance).
- In this norm group, no cut score of 16 or more meets the four-fifths ratio for Black candidates, and at 16 the ratio for women sits on the line. An employer using a

© 2026 Seliant ApS

pass/fail cut should therefore plan a local validation study (Chapter 8) before relying on it.

Outside the US. Equality laws elsewhere also expect a selection method that affects groups differently to be justified by the needs of the job; check local requirements.

Feedback to candidates. Give feedback by band rather than exact score or percentile, and do not show candidates the questions or answers.

© 2026 Seliant ApS

11. Test security and data quality

With two fixed forms, the test stays secure only if retests use the other form and the questions stay confidential.

Two forms. Forms A and B share no questions. A first attempt can be on either form, and the platform puts a retest on the other. A candidate who has taken both forms has seen every question, so a third attempt is not recommended.

Retests. When an employer initiates a retest, the platform automatically assigns the other form, and no minimum waiting period applies. Seliant recommends at most one retest. A retest on the other form raises scores by 0.6 points on average; in the norm group, retaking the same form raised them by 3.2.

Tracking attempts. Track attempts by candidate identity, for example the invitation, not by the candidate's own answer. When asked "Have you taken this test before?", half of the candidates who had in fact taken it answered no.

Integrity signals. On every sitting, the platform records signs that a candidate may be getting outside help:

- switching away from the test tab, or clicking into another app or window
- copying text from the test
- the page becoming noticeably narrower mid-test, as happens when a side panel such as an AI assistant opens (zooming or resizing the window does not count)
- the page using less than 70% of the window at the start, which usually means a side panel was already open; the width and zoom level are recorded, so a panel can be told apart from a zoomed page
- a second screen connected mid-test (detected in Chrome and Edge only)
- leaving full screen (except on phones, the test switches to full screen before it starts, and candidates must stay in it to finish)

Two timing checks run alongside them: correct answers given faster than anyone could read the question, and a whole test finished much faster than usual.

Using the flags. Each signal becomes a "worth a look" flag, never a verdict, and employers see them in a verification note on the candidate's page in the platform; they are not printed on the score report. Flags are common and usually innocent: in the norm group, 24% left the test window at least once, yet they scored no higher than others (0.4 points lower on average) and showed no sign of looking answers up. Treat a flag as a reason to look closer, for example with a supervised retest on the other form, not as grounds for rejection.

Unsupervised testing. The norm data were collected unsupervised online, which is how most employers will use the test. For high-stakes decisions, employers can add online proctoring, in-person supervision or browser lockdown tools, or retest a borderline candidate on the other form under supervision.

© 2026 Seliant ApS

Confidentiality. The test is closed to the public and none of its questions is publicly available; the two practice questions are not used in the test. Employers must not copy, photograph or share the questions, or use them for practice.

Data protection. Candidate responses and scores are stored securely, in compliance with data privacy regulations such as the EU General Data Protection Regulation (GDPR).

© 2026 Seliant ApS

12. Language versions

The test is available in six languages. All versions use the same questions, pictures and answer keys, and all are scored against the English norms.

Language	Variety	Units
English	Original version	pounds and feet
Danish	Standard Danish, informal "du"	kilograms and meters
German	Standard German, formal "Sie"	kilograms and meters
French	Standard French, formal "vous"	kilograms and meters
Spanish	US and Latin American Spanish, "usted"	pounds and feet
Portuguese	Brazilian Portuguese, "você"	pounds and feet

How the versions were made.

- Only the words were translated. Question identifiers, answer keys and the order of the options are unchanged, so every version is scored the same way.
- Wording stays at the same plain, workshop level as the English, using a controlled glossary of about 50 technical terms.
- Terms that would hint at the answer in another language were avoided, and answer options were kept parallel in form so that none stands out.
- Pictures containing words were relabeled in each language; pictures with only letters or symbols are shared by all versions.
- Only unit names change, not the numbers, so every keyed answer stays correct. Spanish and Portuguese keep pounds and feet because most of their candidates are expected to be US-based.
- The instructions and the two practice questions, including their explanations, are fully translated.
- A native speaker of each language reviewed its translation against the English.

Norms and equivalence. All six versions are scored against the English norms in Appendix A. The equivalence of the translations rests on how they were made; there is not yet enough data from the translated versions to compare them statistically with the English one.

Choosing a version. Candidates should take the test in the language they read best. On the English version, candidates whose English is not fluent score up to 2.4 points lower than native speakers (Chapter 9).

© 2026 Seliant ApS

Appendix A: Norm tables

Percentiles against the norm group of 5,460 (Chapter 6). Use the column for the form the candidate took.

Raw score	Percentile, Form A	Percentile, Form B	Band
0–7	<1	<1	Lower
8	1	<1	Lower
9	2	1	Lower
10	3	2	Lower
11	4	4	Lower
12	5	5	Lower
13	7	8	Lower
14	9	11	Lower
15	13	15	Lower
16	17	19	Lower
17	21	24	Lower
18	26	29	Middle
19	32	34	Middle
20	39	40	Middle
21	46	46	Middle
22	54	52	Middle
23	61	58	Middle
24	68	65	Middle
25	74	71	Middle
26	80	78	Upper
27	86	84	Upper
28	91	89	Upper
29	95	93	Upper
30	98	96	Upper
31	99	98	Upper
32	>99	>99	Upper

Percentile = the percentage of the norm group scoring lower, plus half of those with the same score.

© 2026 Seliant ApS

Appendix B: Norm group composition

Shares of the 5,460 results in the norm group, as candidates described themselves.

Characteristic	Breakdown
Sex	Men 71%, women 19%, other or not stated 10%
Age	Under 18: 11%; 18–21: 20%; 22–25: 17%; 26–35: 27%; 36–50: 18%; 51 and over: 7%; median 26
Country	US 48%, Australia 10%, Canada 7%, India 5%, UK 4%, Pakistan 3%, Philippines 3%, other 21%
First language	English 66%; the most common others are Spanish, Urdu, Tagalog, Hindi and Arabic
English level (self-rated)	Fluent 61%, very well 27%, not very well 6%, not at all 1%, not stated 4%
Education	High school graduate 37%, less than 12 years 11%, some college or in college 15%, associate degree 6%, four-year degree 14%, graduate study or degree 7%, other or not stated 9%
Employment	Employed 49%, student 18%, not employed 22%, retired or homemaker 3%, not stated 8%
Job field	Engineering 17%, manufacturing 12%, construction 12%, automotive 10%, admin and clerical 3%, transportation 3%, health care 3%, other 41%
Device	Mobile device, almost all phones, 51%; computer 49%
Form taken	Form A 48%, Form B 52%

Leaving out the 11% under 18, who score lower on average, would lower percentiles by about 1 point.

© 2026 Seliant ApS

Appendix C: Decisions and adverse impact by cut score

A candidate passes with a score at or above the cut. Accuracy figures are simulated from the item response model, except the real-retest column, which uses the 270 candidates who took both forms.

Decision accuracy

Cut	Screened out	Same verdict, simulated retest	Same verdict, real retests	Correct verdict	Wrongly rejected	Wrongly accepted	Pure guessers passing (A / B)
14	8%	93%	91%	95%	2.7%	2.2%	7.2% / 4.8%
15	12%	91%	88%	94%	3.3%	2.8%	3.4% / 2.1%
16	16%	89%	86%	93%	3.8%	3.6%	1.4% / 0.8%
17	20%	88%	84%	91%	4.3%	4.4%	0.5% / 0.3%
18	25%	86%	81%	90%	4.8%	5.3%	0.2% / 0.1%
19	30%	84%	78%	89%	5.1%	6.3%	0.1% / under 0.1%
20	37%	83%	79%	87%	5.4%	7.4%	under 0.1%
21	43%	82%	76%	86%	5.5%	8.4%	under 0.1%

Wrongly rejected: candidates whose true score is at or above the cut but whose observed score falls below it, as a share of all candidates. Wrongly accepted is the reverse.

Adverse impact ratios

Each group's pass rate divided by the comparison group's, in the norm group; ethnicity for US residents only. Ratios below 0.80 fall short of the US four-fifths rule of thumb (Chapter 10).

Cut	Women ÷ men	Black ÷ White	Hispanic ÷ White
14	0.89	0.79	0.96
15	0.85	0.72	0.93
16	0.80	0.68	0.91
17	0.76	0.61	0.90
18	0.73	0.57	0.84
19	0.69	0.51	0.84
20	0.64	0.46	0.80
21	0.58	0.44	0.76

At a cut of 16, the 95% confidence intervals are 0.77–0.84 for women and 0.61–0.74 for Black candidates.

© 2026 Seliant ApS

Appendix D: Item statistics

Both forms have the same average difficulty, but Form A spreads its questions wider, from very easy to very hard, while Form B concentrates them in the middle.

Statistic	Form A	Form B
Questions	32	32
Average proportion correct	0.66	0.66
Range of proportion correct	0.22–0.92	0.37–0.91
Easy questions (80% or more correct)	11	6
Medium questions (50–79% correct)	14	23
Hard questions (under 50% correct)	7	3
Average item-rest correlation	0.30	0.30
Range of item-rest correlation	0.20–0.40	0.15–0.50
Average discrimination (2PL a)	0.93	0.90
Median time per question	19 seconds	18 seconds
Largest misfit to the model (RMSD)	0.06	0.06

The item-rest correlation shows how strongly getting a question right goes with the score on the other 31; the discrimination parameter shows the same from the item response model.

© 2026 Seliant ApS

Appendix E: Glossary

Term	Meaning
Adverse impact	A substantially lower pass rate for one group than for another
Alternate forms	Two versions of a test built to the same specification and difficulty, with different questions
Band	One of three score ranges used for reporting: Lower 0–17, Middle 18–25, Upper 26–32
Browser lockdown tool	Software that stops a candidate from opening other programs, tabs or websites during a test
Confidence interval	A range that contains the true value with a stated probability, usually 95%
Cut score	The lowest score that passes a screen
Decision consistency	The share of candidates who would get the same pass/fail verdict if tested again
Differential item functioning (DIF)	A question being harder for one group than another among candidates of the same overall score
Discrimination	How well a question separates stronger from weaker candidates; in the item response model, the a parameter
Factor	An underlying ability that explains why answers to different questions go together; one dominant factor means the test measures one main ability
Four-fifths rule	The US rule of thumb that treats a pass rate below 80% of the highest group's as evidence of adverse impact
Item response model	A statistical model linking ability to the chance of answering each question correctly; this manual uses the two-parameter logistic (2PL) model
KR-20	A measure of internal consistency for tests scored right or wrong, from 0 to 1
Norm group	The people a candidate's score is compared with
Percentile	The percentage of the norm group scoring lower than a candidate, counting half of those with the same score
Practice effect	The rise in score on a second attempt that comes from familiarity rather than ability
Raw score	The number of correct answers
Reliability	How consistently a test measures, from 0 to 1, where 1 means no measurement error
RMSD	Root mean square deviation: the average gap between observed results and a model's predictions
Split-half reliability	Reliability estimated by splitting the test into two halves and correlating the scores on them
Standard deviation (SD)	The typical spread of scores around the average; group differences are given in SD units
Standard error of measurement (SEM)	The typical size of the error in one person's score

© 2026 Seliant ApS

Term	Meaning
True score	The score a candidate would get on average over many sittings, free of measurement error
Validity	How well the evidence supports the intended use and interpretation of scores

References

- Barrett, G. V., Polomsky, M. D., & McDaniel, M. A. (1999). Selection tests for firefighters: A comprehensive review and meta-analysis. *Journal of Business and Psychology, 13*(4), 507–513.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581–592.
- Muchinsky, P. M. (1993). Validation of intelligence and mechanical aptitude tests in selecting employees for manufacturing jobs. *Journal of Business and Psychology, 7*, 373–382.
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*(2), 297–330.
- Uniform Guidelines on Employee Selection Procedures, 29 CFR 1607.3 (discrimination defined) and 29 CFR 1607.4 (information on impact).
- U.S. Equal Employment Opportunity Commission. Employment tests and selection procedures, including the ban on adjusting scores, using different cut scores or otherwise altering test results by group (42 U.S.C. § 2000e-2(l), added by the Civil Rights Act of 1991).

Version history

Version	Date	Change
1.0	November 2026	First edition, covering Forms A and B and all six language versions

© 2026 Seliant ApS