Psychometrics

BUILDING BETTER ORGANIZATIONS THROUGH PEOPLE
LA DYNAMIQUE INDIVIDUELLE : LA FORCE DE L'ENTREPRISE

ACER Mechanical Reasoning Test

Canadian Technical Supplement

Dr Donald Macnab

Psychometrics Canada

Psychometrics Canada Ltd. · 7125 – 77 Avenue · Edmonton, Alberta · T6B 0B5 · Canada
1-800-661-5158 · **Tel** 780-469-2268 · **Fax** 780-469-2283 · info@psychometrics.com · www.psychometrics.com

Introduction

The ACER Mechanical Reasoning Test is a test of mechanical reasoning ability. This ability can be broadly defined as the ability to perceive and understand relationships between components within a mechanism.

Mechanical Reasoning involves:

- an understanding of the very basic principles of physics in general, and mechanics in particular;
- the ability to visualize movements of objects through space (three dimensional spatial ability); and, most importantly;
- an understanding of the cause-effect relationships between mechanical components

Performance on the ACER Mechanical Reasoning Test is intended to be as independent as possible of mechanical knowledge and information. For this reason, diagrams used in the test show idiosyncratic mechanisms, rather than real-life mechanisms with which some of the candidates may familiar with. For similar reasons, the use of technical terms within the test has been kept to a minimum. Performance on the test is also intended to be independent of reading ability; consequently, the amount of written text within the test has been kept to a minimum.

Uses of the test

There are two main uses of the ACER Mechanical Reasoning Test:

1. Selection of technical and trades personnel, apprentices, trainees, and others involved in the work of a mechanical nature.
2. Counselling of individuals, such as high school students or mature aged job changers, who aspire to occupations requiring mechanical reasoning ability.

As a general rule, the test can be used as part of a broader battery of tests for assessing people's aptitudes for designing, developing, building, repairing or servicing machinery.

It has also been found useful as part of selection and assessment batteries for assessing aspirants to professional careers such as airline pilots and others of an engineering-based nature.

Canadian Reference Sample

This technical supplement contains additional normative data for the ACER Mechanical Reasoning Test (MRT) Form A, derived from a Canadian reference sample of job applicants. The MRT is a 32 item test of mechanical reasoning which has wide ranging content and includes: wheels, gears, clamps, levers, sliding rods, shafts, pulleys, weights, conveyor belts, fixed and non-fixed pivots, and springs.

Sample characteristics

The reference sample comprised 1448 job applicants across a large number of Canadian companies. The sample consisted of 1266 males and 182 females.

Table 1: Applicant performance on the MRT Canadian Sample (n=1448)

	Minimum	Maximum	Mean	Standard Deviation	Percentage Correct
Overall	1	32	16.39	5.31	51%
Males	3	32	16.74	5.24	52%
Females	1	29	13.93	5.12	44%

The finding that males outperform females is in line with virtually all tests of mechanical reasoning ability. The magnitude of the difference in the ACER Mechanical Reasoning test performance levels between males appears to be of a similar order to the differences that have been found in other tests which purport to measure the same aptitude, but are a little less pronounced than in the case of the mechanical reasoning subtest of the widely used Differential Aptitude Test (de Lemos, 1989).

A sub-sample of the applicants provided information about their educational status. Table 2 outlines the descriptive statistics related to this sub-sample.

Table 2: Applicant data by educational level (n=401)

	Minimum	Maximum	Mean	Standard Deviation	Percentage Correct
Some high school (n=147)	4	25	14.56	5.00	46%
High school graduate (n=144)	2	30	16.39	5.58	51%
Trade school (n=63)	7	31	18.16	5.06	57%
University educated (n=47)	1	32	18.57	6.74	58%

ACER Mechanical Reasoning Test Reliability

Reliability is concerned with the consistency of test scores, and how free test results are from external, confounding influences. The higher the reliability of a test, the more likely it is consistently measuring differences between people. More reliable tests provide results that remain unaffected by irrelevant variations, or what is commonly called random errors. Reliability is measured using correlation coefficients. A reliability coefficient is denoted by the letter "r", and is expressed as a number ranging between 0 and 1.00 with r=0 indicating no reliability, and r=1.00 indicating perfect reliability.

It is important to recognize that tests are never 100% accurate, so you will not find a test with a correlation coefficient of r=1.00. In general you will see the reliability of a test expressed as a decimal, for example, r=.80 or r=.93. There are a number of reasons and/or conditions that lead to unreliable test results. Some of the possible reasons include the following.

1. Candidate related. Test performance can be influenced by a person's psychological or physical state at the time of testing. For example, differing levels of anxiety, fatigue, or motivation may affect the individual's test results.

2. Test-related. Item design, instructions, examples and the design of the response procedure can influence an individual's test results. For example, confusing items or complicated instructions which make understanding the test difficult can negatively affect a person's results.

3. Procedural. Differences in the testing environment, such as room temperature, lighting, noise, or even the test administrator and scoring procedures can influence an individual's test performance.

These three factors are sources of chance or random measurement error in the assessment process. If there were no random errors of measurement, the individual would get the same test score, their "true" score, each time. The degree to which test scores are unaffected by measurement errors is an indication of the reliability of the test.

One of the main approaches used to assess reliability is through measures of internal consistency. A sophisticated form of internal consistency reliability is Cronbach's alpha. It effectively splits the test items in every possible way and computes the average of all combinations. Consistency should be achieved such that all the items are measuring the same thing to the same degree, and, therefore, the items for each test scale should have a high degree of correlation. Most professionals agree that test scales with correlation coefficients above .70 are useful for most applications. The internal consistency reliability coefficients for each of the ACER Mechanical Reasoning Test are shown in Table 3. As shown in the table, the reliability coefficients range from .79 to .87 for the whole sample, for males and females and for groups with different educational levels. This exceeds the .70 level for every group, indicating the consistency of MRT test scores is high. The strength of the reliability coefficients indicates that the WPI is relatively free from external errors that could negatively impact the measurement of mechanical reasoning.

Table 3 Reliability Coefficients for various samples of the MRT

Group	N	KR20
Total Sample	1448	.80
Males	1266	.79
Females	182	.83
Some High School	147	.82
High School Graduate	144	.79
Trade School	63	.81
University Educated	43	.87

Score Classification Ranges for the MRT

Below in Table 4 reference group raw scores and percentile ranks are converted into classification ranges commonly applied to measures of ability. While less precise than percentile rankings, these ranges allow performance of any MRT test taker to be classified and discussed in common terms in relation to the reference sample.

Table 4: MRT raw scores and percentile ranks by classification range

Classification Range	Percentile Rank	Stanine	Raw Score
Extremely Low	3 and below	1	0 - 7
Very Low	4-10	2	8 - 9
Below Average	11-22	3	10-11
Slightly Below Average	23-39	4	12-14
Average	40-59	5	15-17
Slightly Above Average	60-76	6	18-19
Above Average	77-89	7	20-23
Well Above Average	90-95	8	24-25
Superior	96 and above	9	26-32

The ACER Mechanical Reasoning Test is available with Canadian applicant norms through careerID.com.

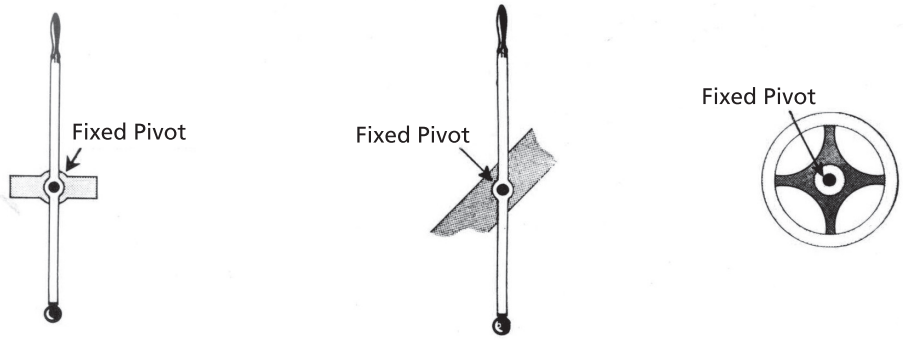
References

Australian Council for Educational Research. (1997) ACER Mechanical Reasoning Test. Melbourne, Australia: ACER Press.

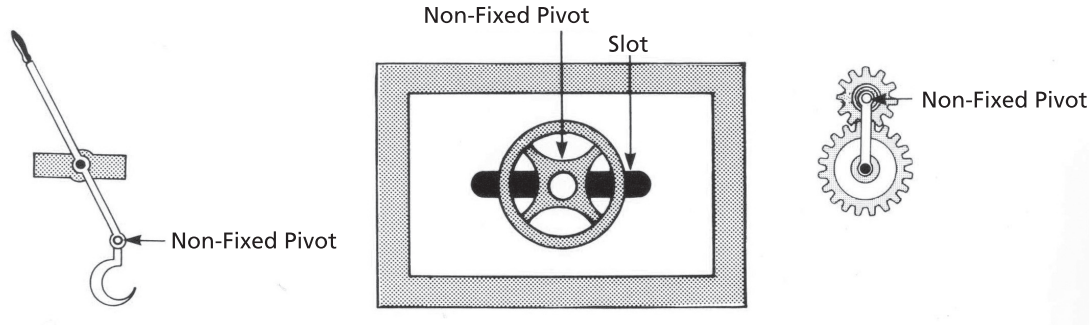
De Lemos, M.M. (1989). Differential Aptitude Tests Form V and W: Australian Manual. Marrickville, NSW: The Psychological Corporation in collaboration with ACER.

ACER Mechanical Reasoning Test
Sample Questions

For this test, you need to understand what a pivot is. A pivot is a pin or axle around which a lever or wheel can tun.
A **fixed pivot** is a pin or axle around which a lever or wheel can turn, but the fixed pivot itself **does not move**. A fixed pivot is indicated by a black dot, as demonstrated below in this example.

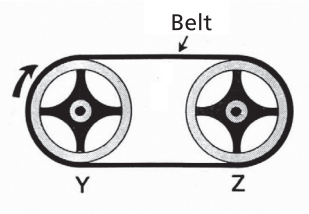


A **non-fixed pivot** is also a pin or axle around which a lever or wheel can turn, but the non-fixed pivot itself **can move**. A non-fixed pivot is indicated by a clear dot, as demonstrated in this example.



SAMPLE 1

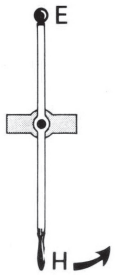
When wheel Y turns clockwise, as shown by the arrow, wheel Z will



A turn clockwise.
B turn anticlockwise.
C turn to and fro.
D stay still.

SAMPLE 2

When handle H is pulled to the right, as shown by the arrow, end E will



A move to the left.
B move to the right.
C move to and fro.
D stay still.

ACER

Mechanical Reasoning - Form A

NAME: John Sample
DATE OF TESTING: 9 Apr 2005

Scores

TOTAL SCORE

The score to the right indicates the number of correct answers this candidate had on this test. The maximum number correct is 32.

21

PERCENTILE

Percentile ranks show the percentage of the reference group with scores below the candidate's score. For example, a percentile rank of 63 means that the candidate has performed better than 63% of the candidates in the reference group; a percentil rank of 25 means that the candidate has performed better than only 25% of the reference group.

84

STANINE

Stanines are standard scores based on a scale of nine equal units that range from a low of 1 to a high of 9. In general, stanines of 1 through 3 are considered below average, 4 through 6 average, and 7 through 9 above average in comparison to the reference group.

7