

TECHNICAL MANUAL & USER GUIDE

GERMAN LANGUAGE TEST

This manual has been created by the psychologists from the Science and Innovation department of Central Test.

The assessment GERMAN LANGUAGE TEST and the logo Central Test are registered trademarks.

No reproduction, modification or adaptation is permitted without the written consent of the publisher Central Test

*Copyright © 2020 by Central Test - Empowering Talent.
All rights reserved.*

Table of Contents

1. Introduction	3
1.1. What does the German Language Test measure?	3
1.2. What does the test taker/test administrator get?	3
2. Factor Descriptions	4
3. Psychometric Properties	5
3.1. Sample Descriptives	5
3.2. Descriptive Statistics	5
3.3. Reliability	7
3.3.1. Internal Reliability	7
3.3.2. Split-half Reliability	7
3.4. Validity	8
3.4.1. Inter-Factor Correlations	8
3.4.2. Confirmatory Factor Analysis (CFA)	9
3.5. Socio-cultural considerations	9
3.5.1. Norms	10
References	12

1. Introduction

Effective Communication skills are an essential requirement for an employee in any organization. An individual who can express his/her views clearly and concisely in written and oral form is considered to be an undoubted asset to the organization. Certain jobs in the print media industry and other profiles such as Sales, Managerial Positions, Content Writing, etc. require superior knowledge of the German language wherein the individual needs to be clear and concise in the way he/she communicates.

Therefore, the **GERMAN LANGUAGE TEST** was created to support companies in recruiting decisions, and especially to examine the German language skills of non-native speakers.

1.1. What does the German Language Test measure?

This test is designed to assess the key aspects of the German language that are most significant and instrumental in a corporate setting which are Vocabulary, Grammar and Reading Comprehension. The focus here is on communication i.e. getting the message across in the appropriate form and receiving messages in their intended sense.

1.2. What does the test taker/test administrator get?

- A summary of candidate's overall performance including number of correct responses and overall language proficiency on a scale of zero to ten.
- Personalized interpretation of overall score.
- Graphical representation of scores on each of the three factors: Reading Comprehension, Vocabulary and Grammar.
- Detailed comments on each factor score.

Suitability of the Test

The **GERMAN LANGUAGE TEST** provides a norm based objective assessment of the test-taker's German language skills. This test is recommended for recruitment of non-native speakers. Due to its level of difficulty, it is less suitable for beginners.

This test is especially reserved for our professional clients like HR departments of organizations, schools and colleges and trainers. This test is not available for general public.

Remote administration of the test

This test is recommended for supervised administration but if the test has to be taken in a non-supervised remote administration mode, appropriate care should be taken by the test administrator to minimize the possibility of getting external help (dictionary, book or expert person) on the part of the test-taker.

Format of the test

- 55 Multiple Choice Questions
- Time: 30 minutes
- Administration: Online

2. Factor Descriptions

The **GERMAN LANGUAGE TEST** is designed in such a way as to measure three important components i.e. Vocabulary, Grammar and Reading Comprehension. These three factors form the crux and prove to be detrimental in language development.

Vocabulary

Vocabulary is a set of words belonging to a specific language. It refers to the words one must know in order to understand and communicate effectively in that language.

Grammar

Grammar is the set of rules governing a particular language such as in word usage, sentence construction etc. Awareness of German grammar is important as it enhances both reading and writing skills, and facilitates effective communication.

Reading Comprehension

Reading comprehension is the level of understanding of a passage or text. It is the capacity of an individual to perceive and understand the meaning communicated in written texts and includes the ability to respond accurately to questions posted from the reading material.

3. Psychometric Properties

The tests at *Central Test* are continuously subjected to fine tuning; this is done in order to ensure long time survival and efficiency of the test. The first validation of the **GERMAN LANGUAGE TEST** was conducted in 2016 on a sample of 102 individuals. The results of this study were as follows.

3.1. Sample Descriptives

The analyses were conducted on a sample of 102 participants between February and July 2016. The tables below display details of the sample characteristics.

Table 1. Age Groups.

Age Group	Percentage
19 – 30 yrs	37%
31 – 45 yrs	35%
46 – 68 yrs	28%

The average age is 38.02 years, with a standard deviation of 12.2 years (minimum age: 19 years, and maximum age: 68 years).

Table 2. Gender.

Gender	Percentage
Male	46%
Female	54%

3.2. Descriptive Statistics

Table 3. Statistical Description of Results for Each Factor.

Factor	Median	Mean	Std. Error	Standard Deviation
Overall	5.50	5.30	0.27	2.70
Reading Comprehension	5.00	4.70	0.26	2.70
Grammar	6.00	5.90	0.29	2.90
Vocabulary	5.80	5.40	0.34	3.40

Mean and standard deviation are important statistics as they help understand a candidate's position within the population of reference.

Standard error of mean provides an estimate of error variance likely to be found in the population. The lesser the standard error in comparison to mean score, the greater the accuracy of the test to assess that particular characteristic.

Mean score is normally a mathematical average of all the observed scores. It provides a tentative assessment as to what extent the data is representative of the theoretical assumptions. Mathematically it tells as to which point on the scale most of the scores are falling around. In this test the factors are scored on a 10-point scale ranging from 0 to 10. Ideally, the mean score should fall between 4.5 and 5.5. However, from the above tabulated scores the mean score of Grammar is falling slightly above the ideal score range.

Standard deviation is the mathematical average of all the individual deviation scores. It provides information as to what extent the scores are dispersed from the mean score.

The table below shows the item sensitivity:

Table 4. Item Sensitivities.

% correct answers		% correct answers	
Q1	51%	Q29	46%
Q2	37%	Q30	71%
Q3	62%	Q31	51%
Q4	48%	Q32	43%
Q5	53%	Q33	73%
Q6	50%	Q34	60%
Q7	40%	Q35	53%
Q8	75%	Q36	65%
Q9	52%	Q37	63%
Q10	38%	Q38	35%
Q11	56%	Q39	58%
Q12	52%	Q40	50%
Q13	63%	Q41	66%
Q14	67%	Q42	71%
Q15	69%	Q43	71%
Q16	72%	Q44	41%
Q17	51%	Q45	45%
Q18	66%	Q46	44%
Q19	67%	Q47	58%
Q20	63%	Q48	56%
Q21	66%	Q49	64%
Q22	40%	Q50	54%
Q23	63%	Q51	51%
Q24	62%	Q52	53%
Q25	57%	Q53	23%
Q26	66%	Q54	50%
Q27	47%	Q55	45%

3.3. Reliability

Reliability measures the extent to which an assessment is free from error and the degree to which items on a scale measure the same construct. For example, all of the items that measure Grammar are coherent and reflect the definition of the factor. A primary concern, when developing and using assessments, is the extent to which random measurement errors influence test performance. If a test is unreliable it is impossible for it to be valid, as accuracy of results (validity) depends on the consistency (reliability) with which the same results are produced. Reliability enables test users to be confident that the results accurately reflect the test taker's attitude, aptitude, or behaviour. In the context of a language assessment, reliability helps assure test users that the scores obtained reflects the language ability of a job candidate or employee. When an assessment has high reliability it also helps to guard against biases that may be caused from errors in human judgement during the recruitment and selection process.

3.3.1. Internal Reliability

Internal, or inter-item, reliability measures the extent to which items on a scale are internally consistent and all measure the same construct. Internal reliability is measured by Cronbach's Alpha (α), which computes the average of all possible inter-item correlations by generating item factor loadings, without relying on a single indicator of reliability which may contain large amounts of error. Cronbach's Alpha ranges from 0 to 1 and has a recommended coefficient of .7, with 1 being total reliability (Nunnally, 1978); values greater than .9 are considered excellent, .7 to .9 is considered good, values of .6 to .7 are acceptable, .5 to .6 is poor, and values of less than .5 is unacceptable.

Table 5. Cronbach's Alpha.

Factor	Reliability
Overall	0.97
Reading Comprehension	0.74
Grammar	0.90
Vocabulary	0.96

Cronbach's Alpha was computed for the overall test as well as for each factor. All factors showed good internal consistency, all alphas being above .74 (see Table). The **GERMAN LANGUAGE TEST** had an overall alpha of .97 indicating that it is very reliable.

3.3.2. Split-half Reliability

Split-half reliability is another measurement of reliability. In split-half reliability, the test is randomly divided and compared; by doing so this helps determine if, and how well, all parts of the test contribute to the measurement of the construct.

An advantage of this approach is that it only requires a single test administration. Split-half reliability

is a correlation between the total scores of the two halves, and so each individual will have two scores, one for each half. Split-half reliability is calculated using the Spearman Brown formula.

The **GERMAN LANGUAGE TEST** scale as a whole appears to have an excellent internal consistency, with a Spearman-Brown coefficient of .936.

3.4. Validity

Validity is the extent to which a test measures what it is supposed to measure and is a judgment made on the basis of observation and statistical indicators. It is important for a test to be reliable at first in order for it to be valid. Validity may be explored and determined through various means:

Face validity

Face validity refers to the operationalization of the concept. It is to see whether "on its face" the test items seem like a good translation of the construct. This is probably the weakest way to try to demonstrate validity as it is highly subjective in nature.

To improve the quality of face validity assessment, and make it more systematic and objective in nature we send the test-items to our panel of experts for their review and on the basis of their feedback items showing consistent agreement are finally retained in the test.

Content validity

Content validity refers to the extent the contents of the test (factor descriptions and items) present true operationalization of the construct. This is again done by the panel of the experts.

Construct validity

Construct validity measures the degree to which a test measures what it claims to measure. Construct validity is essential to the overall validity (perceived and actual) of a scale. When a construct is not clearly defined, direct inferences are unable to be made between construct and other behaviours (Cronbach & Meehl, 1955). Often there is no accepted unit of measurement for constructs and even fairly well-known ones, such as intelligence, are open to debate. The construct validity of a scale may therefore be threatened if a construct is: a) either too broad or too narrow; b) mislabelled as a specific dimension of or component within the construct; c) confounded by other related constructs that mask the effects of the main construct; or d) developed with insufficient research.

Two approaches were used to establish construct validity in the **GERMAN LANGUAGE TEST**: subscale inter-correlations and confirmatory factor analysis.

3.4.1. Inter-Factor Correlations

Correlations were run between the **GERMAN LANGUAGE TEST** factors to explore the relationship between them. Results are shown in the table below.

Table 6. Inter-factor Correlations.

	Reading	Vocabulary	Grammar	Overall
Reading	1.00	0.72	0.70	0.87
Vocabulary	0.72	1.00	0.84	0.94
Grammar	0.70	0.84	1.00	0.93
Overall	0.87	0.94	0.93	1.00

* All correlations are significant at 0.05 level

As illustrated in the table above, there are strong correlations between the **GERMAN LANGUAGE TEST** factors ranging from .70 to .94. Reading, Vocabulary, Grammar and the Overall test outcome all highly correlate with each other. This implies that if a respondent scores highly on one of these factors then it is likely that they will also score highly on the other two, and vice versa.

3.4.2. Confirmatory Factor Analysis (CFA)

To investigate the theoretical constructs or factors that may be represented by a set of items as well as the quality of individual items, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are usually performed. EFA and CFA are both forms of factor analysis. Factor analysis analyses the factors (hypothetical constructs) of an assessment which are developed to interpret clusters of strong inter-correlations between items in a correlation matrix. Factor analysis aims to simplify a range of items into a few common variables; these can be identified by extracting and grouping items which have high correlations on a factor to reveal common variables.

The goal of EFA is to identify and maximise the importance of each dimension (percentage of variance), while CFA is a hypothesis driven approach that determines the degree to which observed indicators reflect the underlying construct by specifying the number of dimensions expected and then classifying the value of expected relationships (Schreiber, Nora, Stage, Barlow, & King, 2006). In doing so, CFA provides a numerical estimate of the degree to which the proposed factor structure is an accurate fit of the phenomenon in the real world. CFA therefore investigated the degree to which measures of a construct are consistent with established understanding and theoretical arguments for that construct. EFA and CFA confirmed that one overall factor was sufficient. All item loadings ranged from 0.50 to 0.94 on that one factor on CFA.

3.5. Socio-cultural considerations

With an objective of examining the socio-cultural influences on the scores of the **GERMAN LANGUAGE TEST**, a comparative statistical analysis was conducted. Comparing test scores among different socio-cultural groups was done for gender and age:

Gender differences

An independent samples t-test on mean was carried out to see whether there is any significant difference between men and women on the dimensions of **GERMAN LANGUAGE TEST**.

Table 7. Independent Samples T-test for Gender.

Factor	Male	Female	p-value t- test
Overall	5.42	5.27	0.79
Reading Comprehension	4.79	4.58	0.70
Grammar	6.00	5.99	0.99
Vocabulary	5.49	5.25	0.72

The means of both men and women were compared to see who performed better on the three factors and the overall test. After comparing the means using the t-test analysis, it was noted that no significant difference was found between men and women on the three factors. This denotes that both men and women performed equally on the **GERMAN LANGUAGE TEST**.

Age effect

Following an inter-factor correlation analysis, no significant correlations between the age of the respondents and their scores on the factors of the **GERMAN LANGUAGE TEST** were found.

3.5.1. Norms

Norms help in the comparison of an individual's score with the scores of others who have completed the same assessment. Norm groups enable us to compare an individual's score to those of a similar population.

Table 8. Factor Score Norms by Gender.

	Mean	Median	Minimum	Maximum	Standard Deviation	Percentile 25	Percentile 75	Percentile 95
Reading Comprehension								
Total	4.70	5.00	0.00	10.00	2.70	2.00	7.00	9.00
Female	4.60	5.00	0.00	10.00	2.80	2.00	7.00	9.00
Male	4.80	4.00	0.00	10.00	2.60	2.00	6.50	9.00
Grammar								
Total	6.00	6.00	0.00	10.00	2.90	4.00	8.40	10.00
Female	6.00	6.50	0.00	10.00	3.20	4.00	9.30	10.00
Male	6.00	6.00	0.00	10.00	2.50	4.30	8.00	9.90

	Mean	Median	Minimum	Maximum	Standard Deviation	Percentile 25	Percentile 75	Percentile 95
Vocabulary								
Total	5.40	6.00	0.00	10.00	3.40	2.80	8.80	10.00
Female	5.30	5.60	0.00	10.00	3.50	2.20	8.80	9.60
Male	5.50	6.40	0.00	10.00	3.40	3.20	8.60	10.00

	Mean	Median	Minimum	Maximum	Standard Deviation	Percentile 25	Percentile 75	Percentile 95
Overall								
Total	5.40	5.45	0.50	9.80	2.70	3.10	7.70	9.50
Female	5.30	5.30	0.50	9.80	2.90	3.00	7.70	9.50
Male	5.40	5.60	0.70	9.70	2.60	3.20	7.70	9.50

Note. Total (n=102), Female (n=55), Male (n=47)

Table 9. Factor Score Norms by German Language Level.

		Beginner (n=12)	Intermediate (n=24)	Fluent (n=34)	Bilingual/Native (n=32)
Reading Comprehension	<i>Mean</i>	3.00	3.21	4.79	6.28
	<i>SD</i>	2.04	2.08	2.40	2.65
	<i>Median</i>	2.50	2.00	5.00	7.00
Grammar	<i>Mean</i>	2.76	5.04	6.16	7.78
	<i>SD</i>	2.15	2.06	2.39	2.82
	<i>Median</i>	3.00	5.00	6.75	9.50
Vocabulary	<i>Mean</i>	1.87	4.13	5.46	7.49
	<i>SD</i>	2.49	2.77	2.85	3.35
	<i>Median</i>	0.40	4.60	6.00	9.20
Overall	<i>Mean</i>	2.52	4.11	5.47	7.18
	<i>SD</i>	1.88	2.02	2.19	2.73
	<i>Median</i>	2.40	3.60	5.65	8.40

References

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. [Http://doi.org/10.1037/h0040957](http://doi.org/10.1037/h0040957)
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, 99(6), 323-338. [Http://doi.org/10.3200/JOER.99.6.323-338](http://doi.org/10.3200/JOER.99.6.323-338)

CONTACT US

Client Service: **customer@centraltest.com**

Training and Consulting: **consulting@centraltest.com**



www.centraltest.com