Gütekriterien und Normierung von TESS

TESS – Test des Englischen für die Sekundarstufe

21.09.2026, Version 7.0.0

Dieses technische Dokument enthält die Gütekriterien und die Normierungsgrundlage von TESS (Test des Englischen für die Sekundarstufe). Es richtet sich an Lehrkräfte und andere Fachkräfte, die Testergebnisse einordnen und interpretieren müssen. Das Dokument wächst mit dem Verfahren: Es wird bei neuen Normierungserhebungen und Erweiterungen fortgeschrieben. Die vollständige Darstellung des Vorgehens und weiterer Kennwerte erfolgt in Fachveröffentlichungen.

Hauptgütekriterien

Die drei Hauptgütekriterien zur Sicherung der Qualität psychologisch-diagnostischer Verfahren sind Objektivität, Reliabilität und Validität.

Objektivität
Objektivität meint die Unabhängigkeit der Testergebnisse von der durchführenden Person als Voraussetzung für die eindeutige Interpretierbarkeit der Ergebnisse. TESS wird digital dargeboten: Instruktionen, Aufgabenreihenfolge und Zeitvorgaben sind für alle Teilnehmenden identisch, die Antworten werden automatisch erfasst und ausgewertet. Auswertungs- und Interpretationsobjektivität sind damit unabhängig von der Person sichergestellt, die den Test anleitet.

Reliabilität
Reliabilität betrifft die Zuverlässigkeit oder Genauigkeit, mit der ein Test die Unterschiede zwischen Personen erfasst. Sie liegt zwischen 0 und 1 und beschreibt das Verhältnis von wahrem Wert und Messfehler: Ein hoher Wert entspricht einem geringen Fehleranteil. Für TESS werden Kennwerte der internen Konsistenz berichtet (Cronbachs α, McDonalds ω) sowie – für die Lesetests – die Split-Half-Reliabilität. Die Werte stehen in Tabelle 2 und Tabelle 3.

Validität
Validität bezieht sich auf die Gültigkeit des Schlusses von den Testergebnissen auf die inhaltlich relevanten Bereiche. Kurz gesagt: Valide ist ein Test, wenn er genau das misst, was er zu messen vorgibt. Sie wird im Fall von TESS über die inhaltliche Repräsentanz der curricularen Anforderungen des Englischunterrichts der Sekundarstufe I und über die theoretische Fundierung der erfassten Teilfertigkeiten sichergestellt. Da mehrere empirische und theoretische Aspekte nötig sind, um Validität zu belegen, werden weitere Aspekte der Validität von TESS in Fachveröffentlichungen adressiert.

© Ernst Klett Verlag GmbH, Rotebühlstraße 77, 70178 Stuttgart 2026. Alle Rechte vorbehalten. www.klett.de
Autoren und Autorin: Dr. Rebecca Silva von Heesen, Dr. Patrick Lösche, Prof. David Gerlach, Prof. Marcus Hasselhorn

Der Test im Überblick

TESS erfasst drei Bereiche des Englischen: Rechtschreibung, Grammatik und Lesen. Aus ihnen werden sechs Indikatoren gebildet, für die Prozentrangnormen vorliegen.

Rechtschreibung (SP)
Die Skala besteht aus 22 Sätzen, in denen Fehler identifiziert werden müssen. Die Sätze bestehen aus insgesamt 94 Wörtern; 31 davon sind fehlerhaft. Die eingebauten Fehler sind authentische Fehler, die Lernenden aus Vorstudien im Diktatformat produzierten. 19 Wörter enthalten einen phonetischen Fehler, 12 Wörter einen Fehler der Groß- und Kleinschreibung. Bei den übrigen 63 fehlerfreien Wörtern wird erfasst, wie häufig fälschlich ein Fehler markiert wird (Falsch-Positiv-Rate).

Der Gesamtwert ergibt sich aus den insgesamt erkannten Fehlern abzüglich der falsch markierten Wörter. Er kann dadurch rechnerisch auch negativ werden. Für die beiden Teilbereiche **phonetische Schreibung (SPP)** und **Groß- und Kleinschreibung (SPC)** liegen zusätzlich eigene Normtabellen vor: Beide lassen sich faktorenanalytisch gut voneinander trennen, und SPC zeigte in den Zusammenhängen mit den übrigen Testteilen ein eigenständiges Profil.

Grammatik (GR)
Die Skala besteht aus 21 Aufgaben, in denen ein Satz mit einem passenden Wort vervollständigt werden muss. Berichtet wird ein Gesamtwert von Inflektionssetzungen bei Verben und Substantiven sowie Homophonwissen.

Lesen (WL, SL, GL)
Der Lesetest besteht aus drei Subskalen: **Wortlesen (WL)**, **Satzlesen (SL) und Textlesen (GL).** Die Leistungen der drei Subskalen werden addiert; daraus werden zwei Kennwerte gebildet:

- **Leseverständnis (LV):** Anzahl der insgesamt korrekt gelösten Aufgaben
- **Lesegeschwindigkeit (LG):** Anzahl der insgesamt bearbeiteten Aufgaben
- **Lesegenauigkeit bzw. Arbeitsstil ergibt sich aus dem Verhältnis von Leseverständnis (LV) zu Lesegeschwindigkeit (LG):** Anteil korrekt gelöster Aufgaben im Verhältnis zu allen bearbeiteten Leseaufgaben, wird als Prozentwert ausgewiesen

Die Aufgaben werden jeweils unter einem Zeitlimit bearbeitet. Daher bearbeiten nicht alle Teilnehmenden gleich viele Aufgaben – genau das macht die Lesegeschwindigkeit interpretierbar.

Indikatoren mit Prozentrangnormen
Rechtschreibung gesamt (SP) · phonetische Schreibung (SPP) · Groß- und Kleinschreibung (SPC) · Grammatik (GR) · Leseverständnis (LV) · Lesegeschwindigkeit (LG)

© Ernst Klett Verlag GmbH, Rotebühlstraße 77, 70178 Stuttgart 2026. Alle Rechte vorbehalten. www.klett.de
Autoren und Autorin: Dr. Rebecca Silva von Heesen, Dr. Patrick Lösche, Prof. David Gerlach, Prof. Marcus Hasselhorn

Die Normierungsstichprobe

Die Normierung beruht auf einer Stichprobe von **N = 1640** Lernenden der Klassenstufen 5 bis 7.

Tabelle 1: Zusammensetzung der Normierungsstichprobe

Merkmal	n	Anteil
Klassenstufe		
Klasse 5	231	14,1%
Klasse 6	747	45,5%
Klasse 7	662	40,4%
Geschlecht		
Weiblich	823	50,2%
Männlich	793	48,4%
keine Angabe	24	1,5%
Gesamt	1640	100%

Anteile gerundet; die Summe kann dadurch von 100 % abweichen.

Reliabilität der Skalen

Tabelle 2 zeigt die interne Konsistenz der Skalen. Berichtet werden Cronbachs α und McDonalds ω. Beide geben an, wie einheitlich die Aufgaben einer Skala dasselbe messen. Zusätzlich ist der Kennwert **Uni** aufgeführt: Er beschreibt, wie stark die Aufgaben einer Skala auf eine einzige gemeinsame Fähigkeit zurückgehen. Für die Rechtschreibung werden die drei Bestandteile getrennt betrachtet und jedes Wort als eigene Aufgabe behandelt.

So lesen Sie die Werte
Alle drei Kennwerte liegen zwischen 0 und 1; höhere Werte sind besser. Für Gruppenvergleiche und die Einordnung einzelner Lernender gelten Werte ab 0,80 als gut.

© Ernst Klett Verlag GmbH, Rotebühlstraße 77, 70178 Stuttgart 2026. Alle Rechte vorbehalten. www.klett.de
Autoren und Autorin: Dr. Rebecca Silva von Heesen, Dr. Patrick Lösche, Prof. David Gerlach, Prof. Marcus Hasselhorn

Tabelle 2: Interne Konsistenz der Skalen (N = 1640)

Merkmal	Items	Cronbachs α	McDonalds ω	Uni
Rechtschreibung (SP)				
Phonetische Schreibung (SPP)	19	0,87	0,87	0,92
Groß- und Kleinschreibung (SPC)	12	0,82	0,84	0,86
Fehlmarkierungen (SP-neg)	63	0,89	0,89	0,65
Fehlmarkierungen gesamt	31	0,89	–	–
Grammatik (GR)	21	0,87	0,88	0,93
Lesen (WL + SL + GL)	3	0,86	0,86	0,98

Items = Zahl der in die Analyse eingehenden Aufgaben bzw. Wörter. **Uni** = Indikator für Eindimensionalität. „Fehlerhafte Wörter gesamt" fasst die 19 phonetischen und die 12 Großschreibungs-Wörter zu einer Skala zusammen.

Die Werte für α und ω liegen durchgehend im guten bis sehr guten Bereich. **SPP** und **GR** zeigen zusätzlich eine hohe Eindimensionalität (Uni = 0,92 bzw. 0,93): Die Aufgaben dieser Skalen gehen erkennbar auf jeweils eine gemeinsame Fähigkeit zurück. **SPC** erreicht mit α = 0,82 und Uni = 0,86 ebenfalls eine gute interne Konsistenz bei klarer gemeinsamer Dimension.

SP-neg erfasst Fehlmarkierungen an 63 fehlerfreien Wörtern – also Fälle, in denen Lernende irrtümlich einen Fehler markieren. Die Skala weist zwar eine hohe interne Konsistenz auf (α = 0,89), aber keine eindeutige Eindimensionalität (Uni = 0,65). Der Grund ist ein Deckeneffekt: Fälschlich markierte Wörter kommen selten vor. Dadurch ist die Streuung der Werte eingeschränkt, was die niedrigen Zusammenhänge zwischen den einzelnen Wörtern erklärt. Für die Praxis bedeutet das: Der Wert ist als Korrektur im Gesamtscore aussagekräftig, nicht als eigenständige Skala.

Reliabilität der Lesetests

Die Lesetests werden unter Zeitlimit bearbeitet, und die Lösungsraten sind hoch. Unter diesen Bedingungen sind übliche Schätzungen der internen Konsistenz nicht sinnvoll: Wer eine Aufgabe nicht erreicht hat, hat sie nicht etwa falsch gelöst. Deshalb wird die **Split-Half-Reliabilität** berichtet: Aus den geraden und den ungeraden Aufgaben wird je ein Summenwert gebildet, deren Zusammenhang berechnet und anschließend nach der Spearman-Brown-Formel korrigiert.

Tabelle 3: Split-Half-Reliabilität der Lesetests

Subskala	Lösungsrate	Split-Half (Spearman-Brown)
Wortlesen (WL)	96%	0,97
Satzlesen (SL)	86%	0,95
Textlesen (GL)	69%	0,88

Die sehr hohen Werte für WL und SL sind eine Folge der niedrigen Fehlerraten. Der Wert für das Textlesen (GL) ist der aussagekräftigste, weil hier genügend Varianz in den Lösungen entsteht.

Zusätzlich wurde die Gesamtskala Lesen betrachtet, indem die drei Subskalen als Aufgaben aufgefasst wurden (α = 0,86; ω = 0,86; Uni = 0,98). Der sehr hohe Uni-Wert spricht für einen starken gemeinsamen Lesefaktor.

Klett © Ernst Klett Verlag GmbH, Rotebühlstraße 77, 70178 Stuttgart 2026. Alle Rechte vorbehalten. www.klett.de
Autoren und Autorin: Dr. Rebecca Silva von Heesen, Dr. Patrick Lösche, Prof. David Gerlach, Prof. Marcus Hasselhorn

Zusammenhänge der Teilbereiche der Rechtschreibung

Die Korrelationen zeigen, wie eng die drei Bestandteile der Rechtschreibung zusammenhängen. Der Korrelationskoeffizient (r) liegt zwischen −1 und +1. Eine positive Korrelation lässt sich lesen als „je höher Wert A, desto höher Wert B“. Ein kleiner Zusammenhang liegt bei etwa r ≈ 0,1, ein mittlerer bei r ≈ 0,3; Zusammenhänge über r ≈ 0,5 können als große Effekte interpretiert werden (vgl. z. B. Döring & Bortz, 2016).

Tabelle 4: Korrelationen der Teilbereiche der Rechtschreibung (N = 1640)

Subskala	SPP	SPC	SP-neg
Wortlesen (WL)	–		
Satzlesen (SL)	0,41	–	
Textlesen (GL)	-0,57	-0,30	–

Alle Zusammenhänge sind statistisch bedeutsam. SP-neg zählt die fälschlich als fehlerhaft markierten Wörter – ein **hoher** Wert steht hier für **mehr** Fehlmarkierungen. Die negativen Vorzeichen sind daher inhaltlich erwartungsgemäß.

Der mittlere Zusammenhang zwischen SPP und SPC (r = 0,41) zeigt: Beide Teilbereiche hängen zusammen, messen aber nicht dasselbe. Das begründet die getrennten Normtabellen. Wer viele Fehler übersieht, markiert zugleich seltener fälschlich fehlerfreie Wörter (r = −0,57 zwischen SPP und SP-neg) – der Gesamtwert SP verrechnet beide Tendenzen miteinander.

Deskriptive Kennwerte

Tabelle 5 gibt die Verteilung der Testwerte in der Normierungsstichprobe wieder. Diese Werte sind keine Normen: Für die Einordnung eines einzelnen Lernenden sind die Prozentrangnormen der jeweiligen Klassenstufe maßgeblich. Die Tabelle zeigt, in welchem Bereich sich Testwerte üblicherweise bewegen.

Tabelle 5: Deskriptive Kennwerte der Gesamtstichprobe (N = 1640)

Kennwert	möglicher Wertebereich	M	SD	Median	Min	Max
Rechtschreibung						
Gesamtwert (SP)	-63 - 31	13,11	9,66	14	-41	31
Phonetische Schreibung (SPP)	0 - 19	11,41	4,77	12	0	19
Groß-/Kleinschreibung (SPC)	0 - 12	4,89	3,23	5	0	12
Grammatik						
Gesamtwert (GR)	0 - 21	14,18	5,04	15	1	21
Lesen						
Wortlesen (WL)	–	24,50	6,18	24	0	46
Satzlesen (SL)	–	20,36	6,50	21	0	38
Textlesen (GL)	–	10,72	4,41	11	0	19
Leseverständnis (LV)	–	55,58	15,08	55,5	12	101
Lesegeschwindigkeit (LG)	–	61,95	12,21	61	22	103
Lesegenauigkeit (in %)	0 - 100	88,86	12,12	93	24	100

M = Mittelwert, **SD** = Standardabweichung. Der Gesamtwert der Rechtschreibung (SP) kann negativ werden, weil falsch markierte Wörter ab-gezogen werden. WL, SL, GL, LV und LG haben wegen des Zeitlimits keine feste Obergrenze.

© Ernst Klett Verlag GmbH, Rotebühlstraße 77, 70178 Stuttgart 2026. Alle Rechte vorbehalten. www.klett.de
Autoren und Autorin: Dr. Rebecca Silva von Heesen, Dr. Patrick Lösche, Prof. David Gerlach, Prof. Marcus Hasselhorn

Kennwerte nach Klassenstufe

Die Leistungen steigen über die Klassenstufen hinweg an – in der Rechtschreibung, in der Grammatik und in allen drei Lesetests. Dieser Anstieg ist der Grund, warum für jede Klassenstufe eigene Prozentrangnormen vorliegen: Ein identischer Rohwert bedeutet in Klasse 5 etwas anderes als in Klasse 7.

Tabelle 6: Mittelwerte und Standardabweichungen nach Klassenstufe

	Klasse 5 (n = 231)		Klasse 6 (n = 747)		Klasse 7 (n = 662)	
Kennwert	M	SD	M	SD	M	SD
Gesamtwert Rechtschreibung (SP)	9,93	8,60	13,08	9,85	14,24	9,56
Phonetische Schreibung (SPP)	9,30	4,42	11,20	4,75	12,38	4,64
Groß-/Kleinschreibung (SPC)	4,63	2,91	5,09	3,33	4,75	3,22
Grammatik	12,35	4,89	14,09	4,87	14,92	5,12
Wortlesen (WL)	22,75	5,74	24,19	6,14	25,47	6,21
Satzlesen (SL)	17,79	5,48	19,78	6,42	21,91	6,54
Textlesen (GL)	8,96	4,07	10,27	4,33	11,83	4,34

Die Groß- und Kleinschreibung (SPC) ist der einzige Bereich ohne durchgängigen Anstieg über die Klassenstufen. Sie ist damit ein Hinweis darauf, dass orthografische Regelkenntnis im Englischen nicht automatisch mit der Verweildauer im Unterricht wächst.

Hinweis zur Interpretation

Mittelwert und Standardabweichung beschreiben die Normierungsstichprobe, nicht die einzelne Lerngruppe. Ein Wert unterhalb des Mittelwerts ist kein Befund – rund die Hälfte aller Lernenden liegt definitionsgemäß darunter. Erst der Prozentrang macht aus einem Rohwert eine einordnende Aussage.

Klett © Ernst Klett Verlag GmbH, Rotebühlstraße 77, 70178 Stuttgart 2026. Alle Rechte vorbehalten. www.klett.de
Autoren und Autorin: Dr. Rebecca Silva von Heesen, Dr. Patrick Lösche, Prof. David Gerlach, Prof. Marcus Hasselhorn

Anhang: erweiterte Reliabilitätskennwerte

Die folgenden Kennwerte stammen aus der Reliabilitätsanalyse des R-Pakets **psych** und ergänzen Tabelle 2. Sie sind für die Interpretation einzelner Testergebnisse nicht erforderlich.

Tabelle 7: Ergänzende Reliabilitätskennwerte (N = 1640)

Skala	Items	ω_h	ω_{tot}	Uni	ECV	mittleres r
Phonetische Schreibung (SPP)	19	0,74	0,87	0,92	0,92	0,26
Groß-/Kleinschreibung (SPC)	12	0,61	0,84	0,86	0,83	0,28
Fehlerfreie Wörter (SP-neg)	63	0,57	0,89	0,65	0,83	0,11
Grammatik (GR)	21	0,74	0,88	0,93	0,92	0,25
Lesen (WL + SL + GL)	3	0,86	0,86	0,98	0,68	0,66

ω_h = hierarchisches Omega (Anteil der Varianz, der auf einen Generalfaktor zurückgeht). ω_{tot} = Omega total. Uni = Indikator für Eindimensionalität. **ECV** = durch den gemeinsamen Faktor erklärter Varianzanteil.
mittleres r = durchschnittliche Korrelation zwischen den Aufgaben einer Skala.

Die ECV-Werte liegen zwischen 0,68 und 0,92. Der niedrigste Wert entfällt auf die Gesamtskala Lesen – dort gehen nur drei Subskalen als Aufgaben in die Analyse ein, was den Kennwert systematisch dämpft.

Zusammenfassung

Für sechs Indikatoren aus drei Testbereichen liegen Prozentrangnormen vor. Die interne Konsistenz der Skalen liegt insgesamt in einem sehr guten Bereich: Cronbachs α liegt je nach Skala zwischen 0,82 und 0,89, McDonalds ω zwischen 0,84 und 0,89. Für die Lesetests liegt die Split-Half-Reliabilität zwischen 0,88 und 0,97. Zusammen mit den Kennwerten zur Varianzaufklärung durch den gemeinsamen Faktor (ECV = 0,68 bis 0,92) sprechen diese Ergebnisse für konsistente und zuverlässige Skalen.
Die Normierung beruht auf einer Stichprobe von N = 1640 Lernenden der Klassenstufen 5 bis 7.

© Ernst Klett Verlag GmbH, Rotebühlstraße 77, 70178 Stuttgart 2026. Alle Rechte vorbehalten. www.klett.de
Autoren und Autorin: Dr. Rebecca Silva von Heesen, Dr. Patrick Lösche, Prof. David Gerlach, Prof. Marcus Hasselhorn