6. Bremer Symposion zum Sprachenlernen und -lehren

Bremen, 24. – 25. Februar 2017

Lücken schließen, Brücken bauen: Bestimmung von GER-Niveaus mit dem onSET

Thomas Eckes

Gesellschaft für Akademische Studienvorbereitung und Testentwicklung (g.a.s.t. e.V.) / TestDaF-Institut
Ruhr-Universität Bochum

25. Februar 2017

g.a.s.t.

onSET
online-Spracheinstufungstest

Übersicht

1. Konzeption des onSET
2. onSET-Teilnehmerergebnisse
3. Brückenschlag zum GER
4. Prototypgruppenmethode (PGM)
5. Bestimmung der Cut-Scores (GER-Niveaus)
6. Zusammenfassung

g.a.s.t.

onSET
online-Spracheinstufungstest

Konzeption des onSET

- onSET = Online-Spracheinstufungstest (www.onset.de)
- Komplett internetgestütztes System zur Messung der allgemeinen Sprachkompetenz in Fremdsprachen
- C-Test-Prinzip (Prinzip der reduzierten Redundanz)
- Aktuell verfügbar: onSET-Deutsch (vormals onDaF), onSET-English

g.a.s.t.

onSET
online-Spracheinstufungstest

Konzeption des onSET (Forts.)

- „Lücken schließen“: Fehlende Wortteile ergänzen (zentral: lexiko-grammatikalische Kompetenz)
- onSET-Methodik: Rasch-Skalierung
  - Systematische Erprobung und Skalierung von Lückentexten
  - Kalibrierte Itembank („Herzstück“)
  - LOFT-Methode der Testdarbietung
  - Acht Texte mit je 20 Lücken

Konzeption des onSET (Forts.)

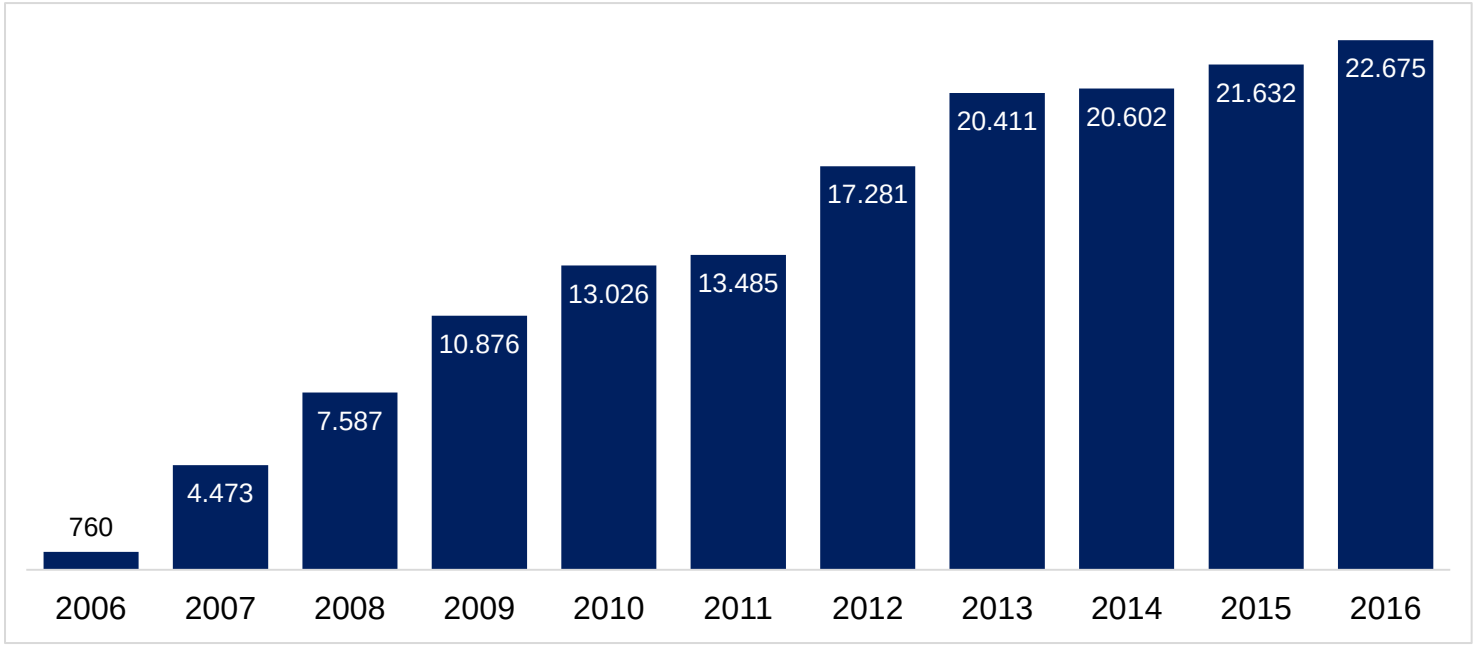
- „Brücken bauen“: Von der Punktzahl zu GER-Niveaus
- onSET-Skala
  - Punktzahl: 0 bis 160 Punkte
  - GER-Niveaus: A2 bis C1

g.a.s.t.

onSET
online-Spracheinstufungstest

onSET-Teilnahme

Bis Ende 2016 haben insgesamt **152.808** Personen den onSET abgelegt.

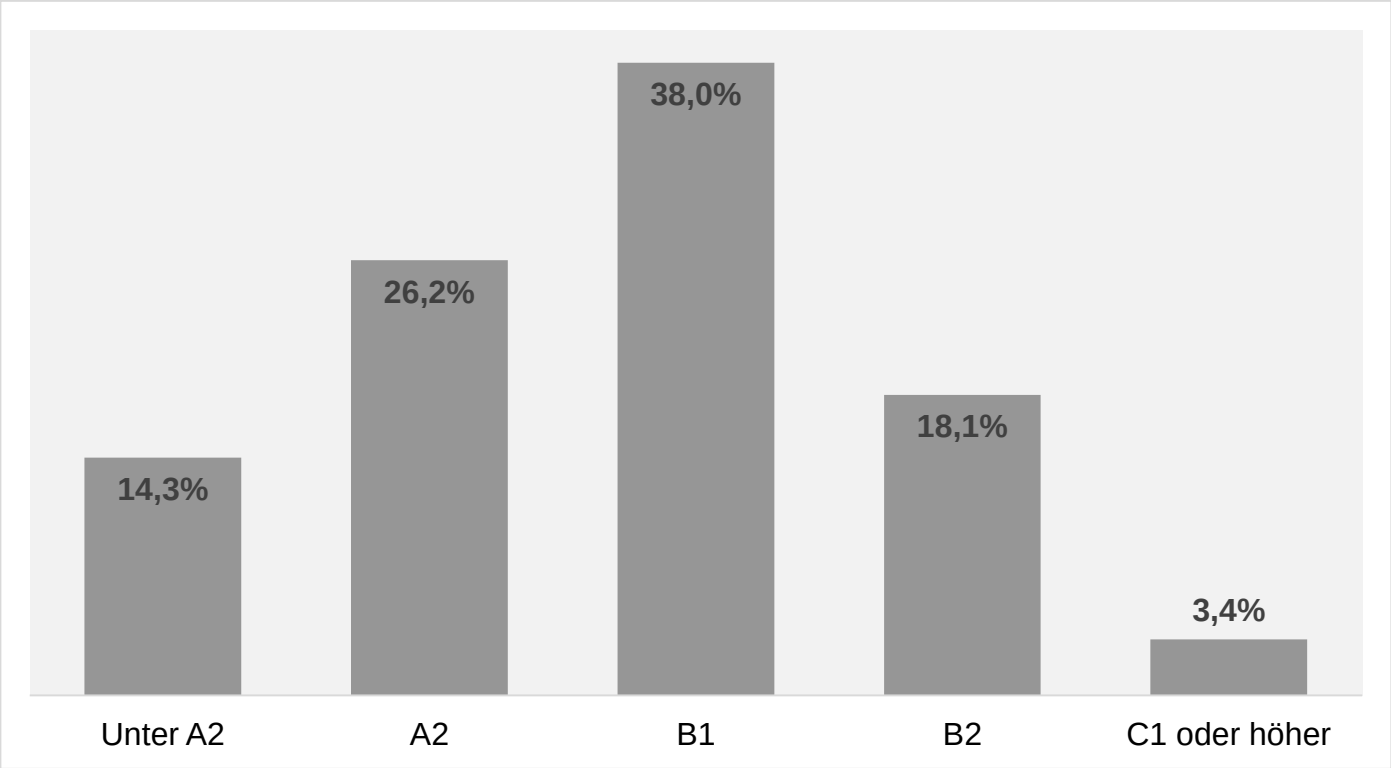


g.a.s.t.

onSET
online-Spracheinstufungstest

onSET-Ergebnisse

Im onSET-Deutsch erreichten 59,5 % der Teilnehmenden das Niveau B1 oder höher.

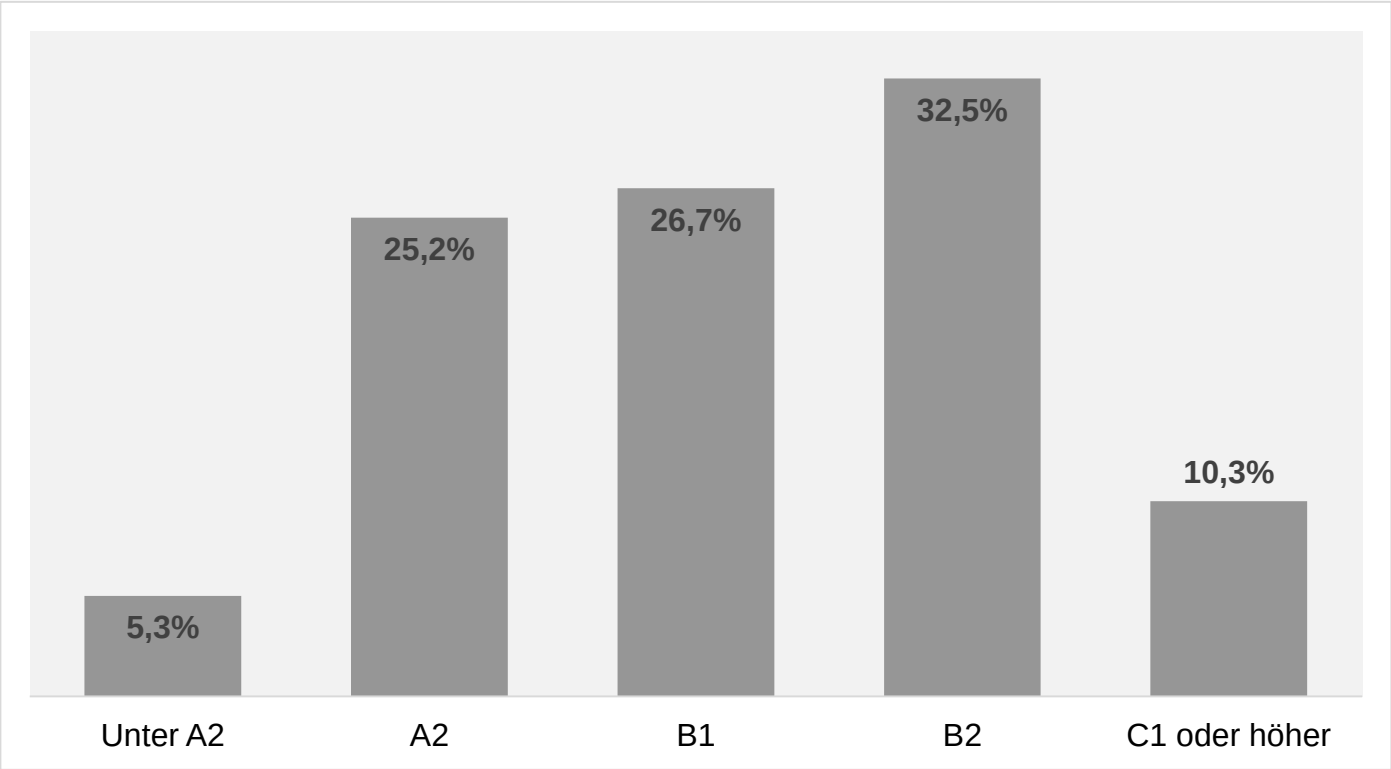


g.a.s.t.

onSET
online-Spracheinstufungstest

onSET-Ergebnisse

Im onSET-English erreichten 69,5 % der Teilnehmenden (*N* = 3.401) das Niveau B1 oder höher.



g.a.s.t.

onSET
online-Spracheinstufungstest

Wie aber erfolgt die Zuordnung von C-Test-Leistungen zu GER-Niveaus?

g.a.s.t.

onSET
online-Spracheinstufungstest

g.a.s.t.

onSET
online-Spracheinstufungstest

- Allgemeine Sprachkompetenz
- Textschwierigkeit komplex

↔

- Kommunikative Kompetenz
- Handlungsorientierung

g.a.s.t.

onSET
online-Spracheinstufungstest

„Unknown territory“?

(Reichert et al., 2010, S. 206)

g.a.s.t.

onSET
online-Spracheinstufungstest

Empirische Stützung für den Brückenschlag zum GER:

1. Korrelation von C-Tests mit rezeptiven und produktiven Fertigkeiten
2. Standard-Setting mit der Prototypgruppenmethode (PGM)

g.a.s.t.

onSET
online-Spracheinstufungstest

Stützpfeiler 1: Korrelationen (und mehr)

Korrelationen zwischen deutschen C-Tests und TestDaF-Subtests (Eckes, 2014; Eckes & Grotjahn, 2006)

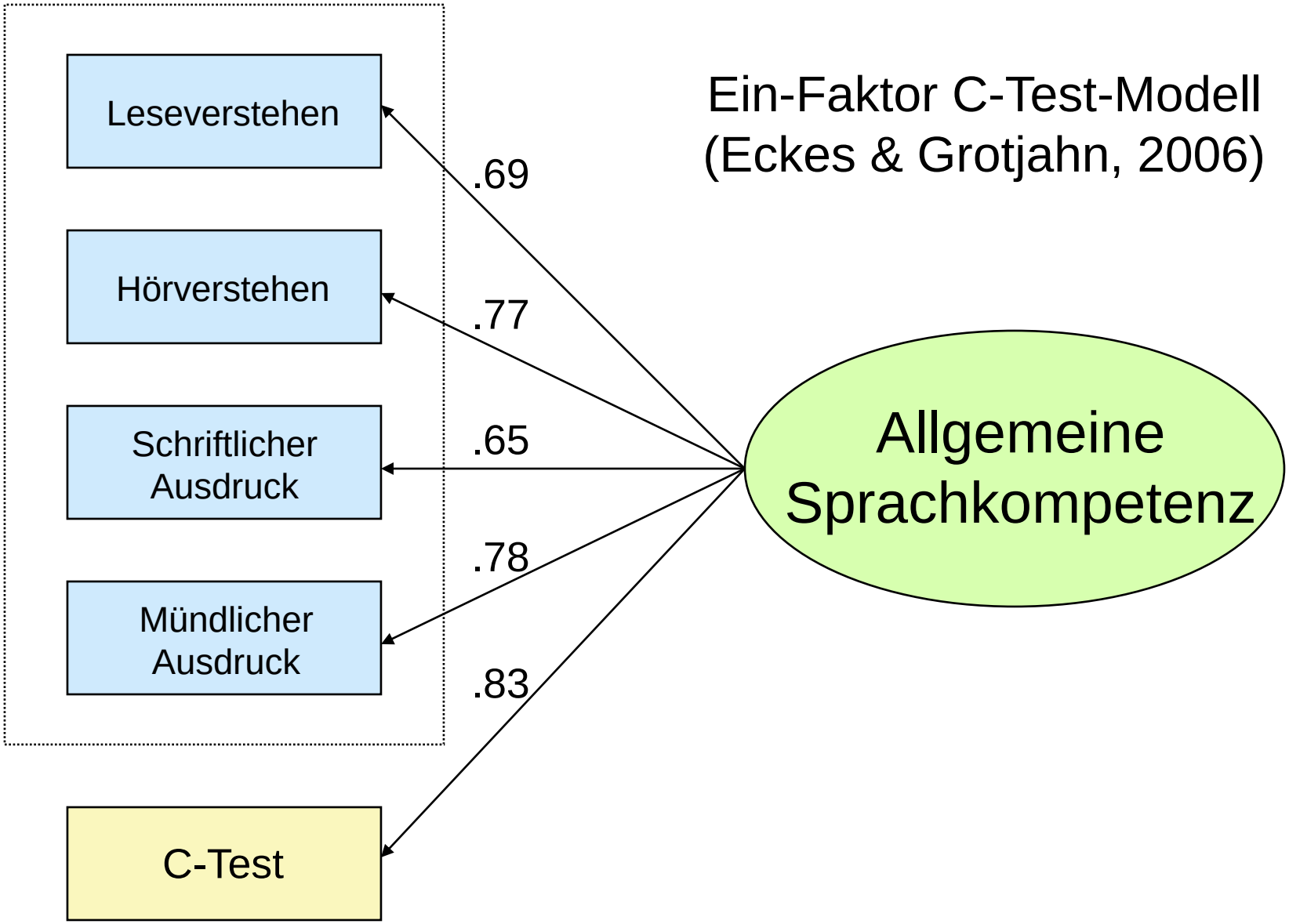
TestDaF-Subtest	Korrelationen
Leseverstehen (Punktzahl, TDN)	.60 bis .66
Hörverstehen (Punktzahl, TDN)	.62 bis .69
Schriftlicher Ausdruck (TDN)	.62 bis .68
Mündlicher Ausdruck (TDN)	.54 bis .64

Alle Korrelationen: $p < .01$.

g.a.s.t.

onSET
online-Spracheinstufungstest

Ein-Faktor C-Test-Modell (Eckes & Grotjahn, 2006)



g.a.s.t.

onSET online-Spracheinstufungstest

Korrelationen eines C-Tests mit Tests des Lese- und Hörverstehens und einem Wortschatztest (Harsch & Hartig, 2016)

Test	LV	HV	W-HR	W-FAR
C-Test	**.73*** **	**.76*** **	.48***	-.15*
Leseverstehen	-	.68***	.39**	-.16**
Hörverstehen	-	-	.49***	-.10
Wortschatz (HR)	-	-	-	.33***
Wortschatz (FAR)	-	-	-	-

W-HR = Wortschatztest (X-Lex) Hit Rate. W-FAR = Wortschatztest (X-Lex) False-Alarm Rate (Guessing). * $p < .05$. ** $p < .01$. *** $p < .001$.

g.a.s.t.

onSET online-Spracheinstufungstest

Vorhersagekraft von C-Test und Wortschatztest für rezeptive Fertigkeiten (Harsch & Hartig, 2016)

Prädiktoren	**Kriterium LV**	**Kriterium HV**
C-Test (allein)	**.53*** **	**.58*** **
Wortschatztest (HR, FAR)	.24***	.32***
Alle (C-Test, HR, FAR)	**.54*** **	**.60*** **

Werte sind Maße der Varianzaufklärung (quadrierter Determinationskoeffizient).
*** $p < .001$.

g.a.s.t.

onSET
online-Spracheinstufungstest

Stützpfeiler 2: Standard-Setting

- onSET-Testwerte: 0 bis 160 Punkte
- GER-Niveaus: A2, B1, B2, C1

- Ab welchem Testwert (x) hat eine Person das Niveau A2, B1, B2 oder C1 erreicht?

g.a.s.t.

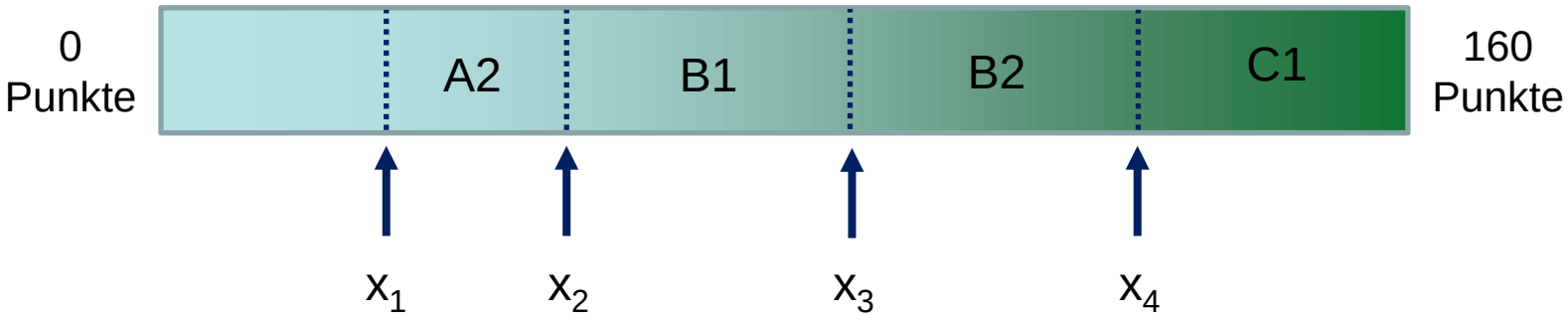
onSET
online-Spracheinstufungstest

Wo liegen die Grenzen zwischen benachbarten Niveaus?

g.a.s.t.

onSET
online-Spracheinstufungstest

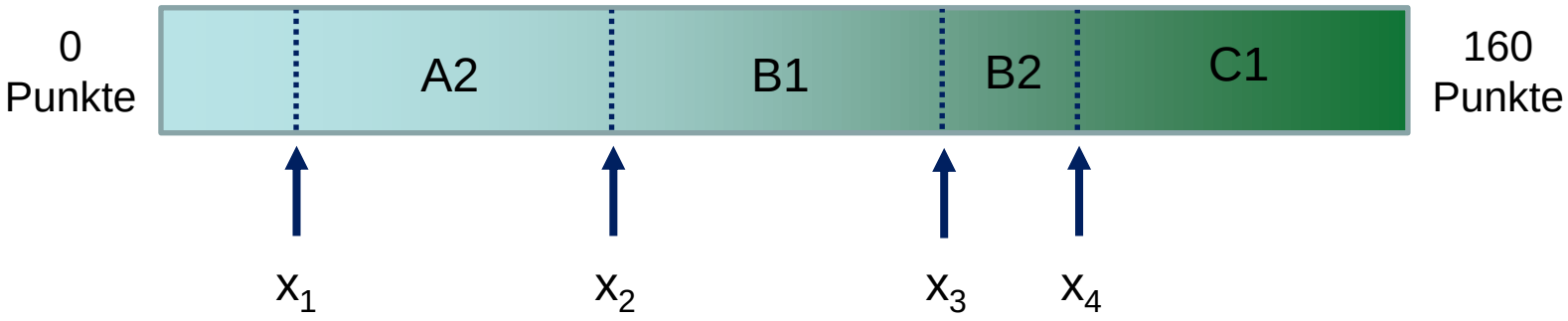
Hier?



g.a.s.t.

onSET
online-Spracheinstufungstest

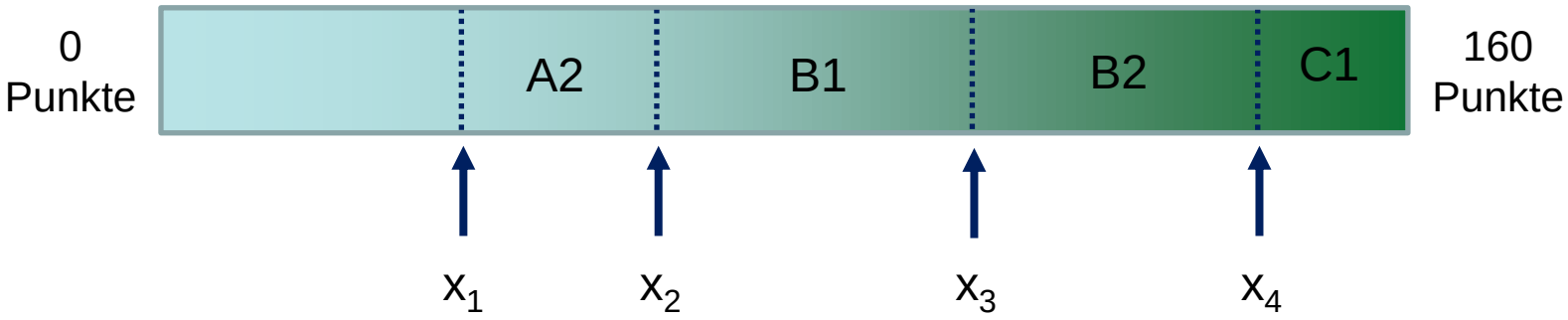
Oder hier?



g.a.s.t.

onSET
online-Spracheinstufungstest

Oder doch eher hier?



g.a.s.t.

onSET
online-Spracheinstufungstest

Methoden des Standard-Settings

- Standard-Setting stützt sich wesentlich auf Urteile oder Einschätzungen geschulter Beurteiler (Experten)
- Hambleton & Pitoniak (2006, p. 235): “blend of judgment, psychometrics, and practicality”
- Testzentriert: Urteile über Testitems
  - Angoff-Methode
  - Bookmarkmethode
- Personenzentriert: Urteile über Teilnehmende
  - Borderlinegruppenmethode
  - Kontrastgruppenmethode

Probleme

- Wahrscheinlichkeitsurteile stark fehleranfällig (Heuristiken, Urteilstendenzen, mangelnde Übereinstimmung zwischen Experten)
- Konzept einer “Borderline-” oder “mindest-kompetenten Person” unklar bzw. diffus
- Testzentrierte Methoden beim onSET nicht praktikabel (ca. 15 bis 20 Experten diskutieren in einem mehrtägigen Workshop mehrere hundert Lückentexte bzw. ein paar tausend Lücken)

Lösung

- **Personenzentrierter Ansatz** (Personen mit bekanntem Sprachleistungsprofil)
- Beurteilung nicht der Grenzfälle, sondern der **typischen** Fälle: **Prototypen** (Konzept aus der Begriffs- und Kategorisierungsforschung)
- Benennung von Prototypen durch GER-erfahrene Sprachlehrkräfte bzw. Kursleiter (im Rahmen von Erprobungen)

Lösung

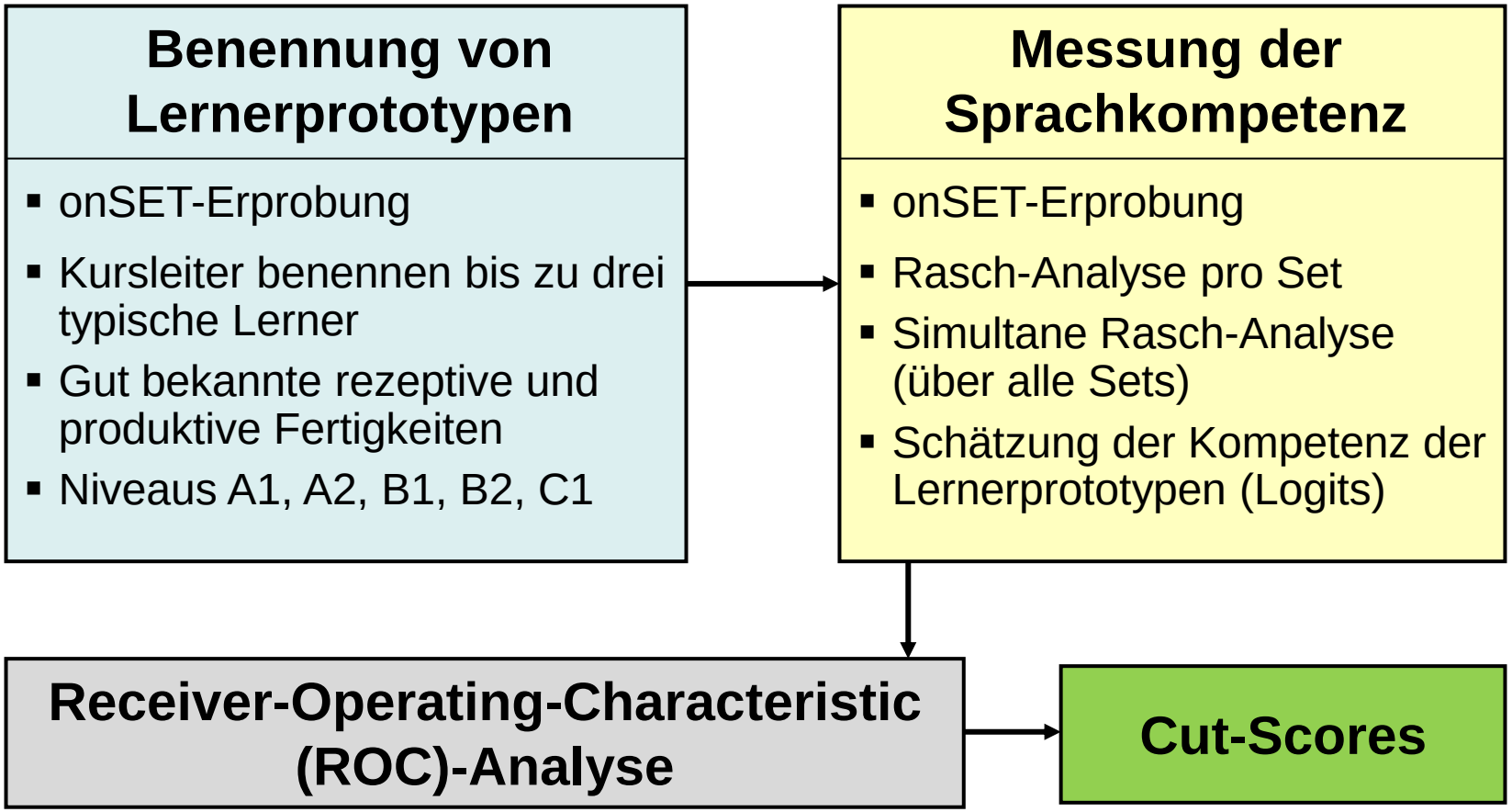
- **Personenzentrierter Ansatz** (Personen mit bekanntem Sprachleistungsprofil)
- Beurteilung nicht der Grenzfälle, sondern der **typischen** Fälle: **Prototypen** (Konzept aus der Begriffs- und Kategorisierungsforschung)
- Benennung von Prototypen durch GER-erfahrene Sprachlehrkräfte bzw. Kursleiter (im Rahmen von Erprobungen)

Prototypgruppenmethode (PGM; Eckes, 2010, 2012, 2016)

g.a.s.t.

onSET online-Spracheinstufungstest

Prototypgruppenmethode



g.a.s.t.

onSET
online-Spracheinstufungstest

Prototypgruppenmethode

ROC-Analyse

- Weit verbreitet in der medizinischen Diagnostik (klinische Testdaten zur Unterscheidung zwischen “gesund” und “krank”)
- Verteilungsfreies, nichtparametrisches Verfahren
- Geringe Anforderungen an Umfang der Stichprobe
- Im Kontext von Methoden des Standard-Settings erstmals diskutiert von Kaftandjieva (2010)

ROC-Analyse: Vierfelder-Klassifikationsschema

Beispiel für Niveau A2 (-) vs. B1 (+)

	Klassifikation (nach Cut-Score)	
Kriterium (B1)	A2 (-)	B1 (+)
Prototyp A2 (-)	RN	FP
Prototyp B1 (+)	FN	RP

+ bedeutet: erfüllt das Kriterium; - bedeutet: erfüllt das Kriterium nicht

RP = richtig-positiv (Treffer)
FN = falsch-negativ (Verpasser)
FP = falsch-positiv (falscher Alarm)
RN = richtig-negativ (korrekte Ablehnung)

ROC-Analyse: Sensitivität und Spezifität

Beispiel für Niveau A2 (-) vs. B1 (+)

	Klassifikation (nach Cut-Score)	
Kriterium (B1)	A2 (-)	B1 (+)
Prototyp A2 (-)	**RN**	FP
Prototyp B1 (+)	FN	**RP**

Sensitivität = RP/(FN + RP)
1 – Sensitivität = FN/(FN + RP)

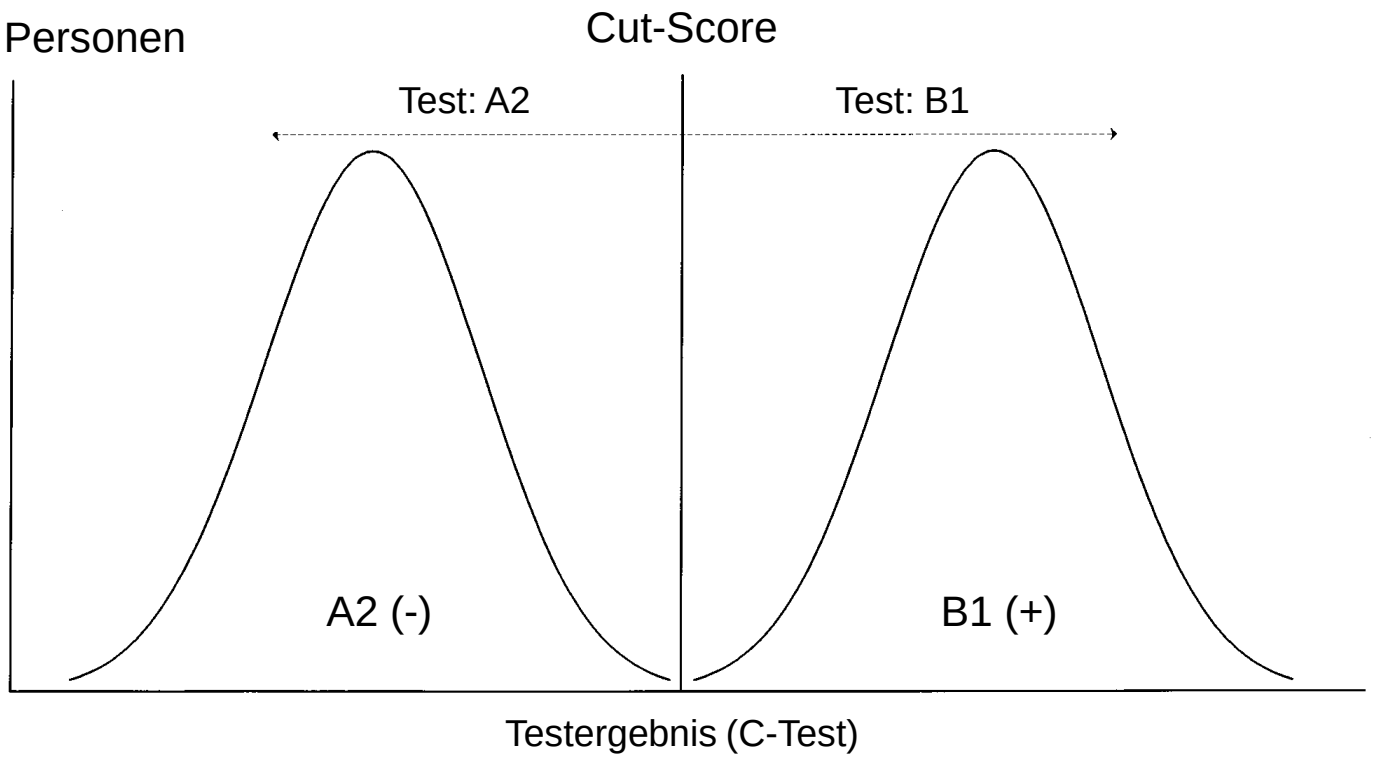
Trefferquote
Verpasserquote

Spezifität = RN/(FP + RN)
1 – Spezifität = FP/(FP + RN)

Quote korrekter Ablehnungen
Quote falscher Alarme

Prototypgruppenmethode

ROC-Analyse: Der Idealfall

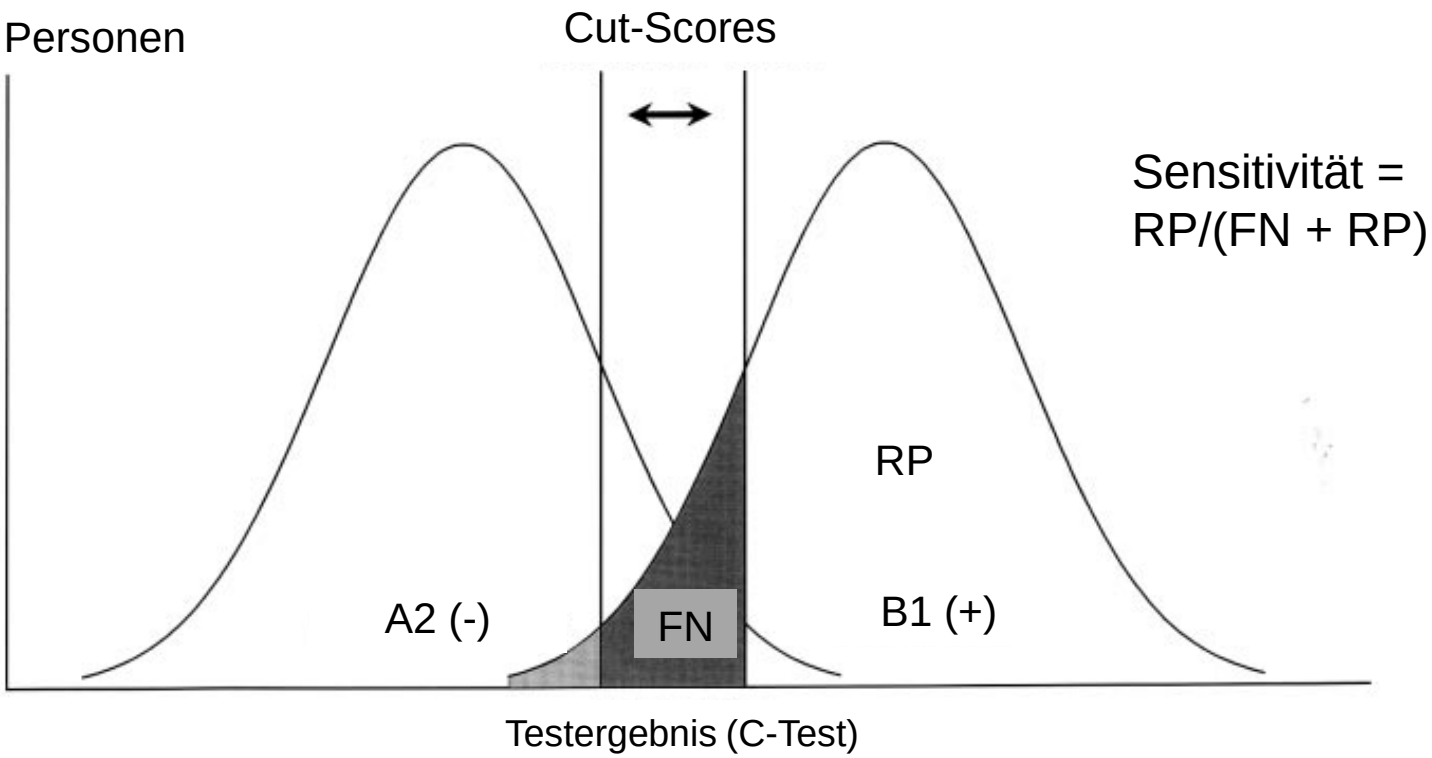


g.a.s.t.

onSET
online-Spracheinstufungstest

Prototypgruppenmethode

ROC-Analyse: Sensitivität (Trefferquote)

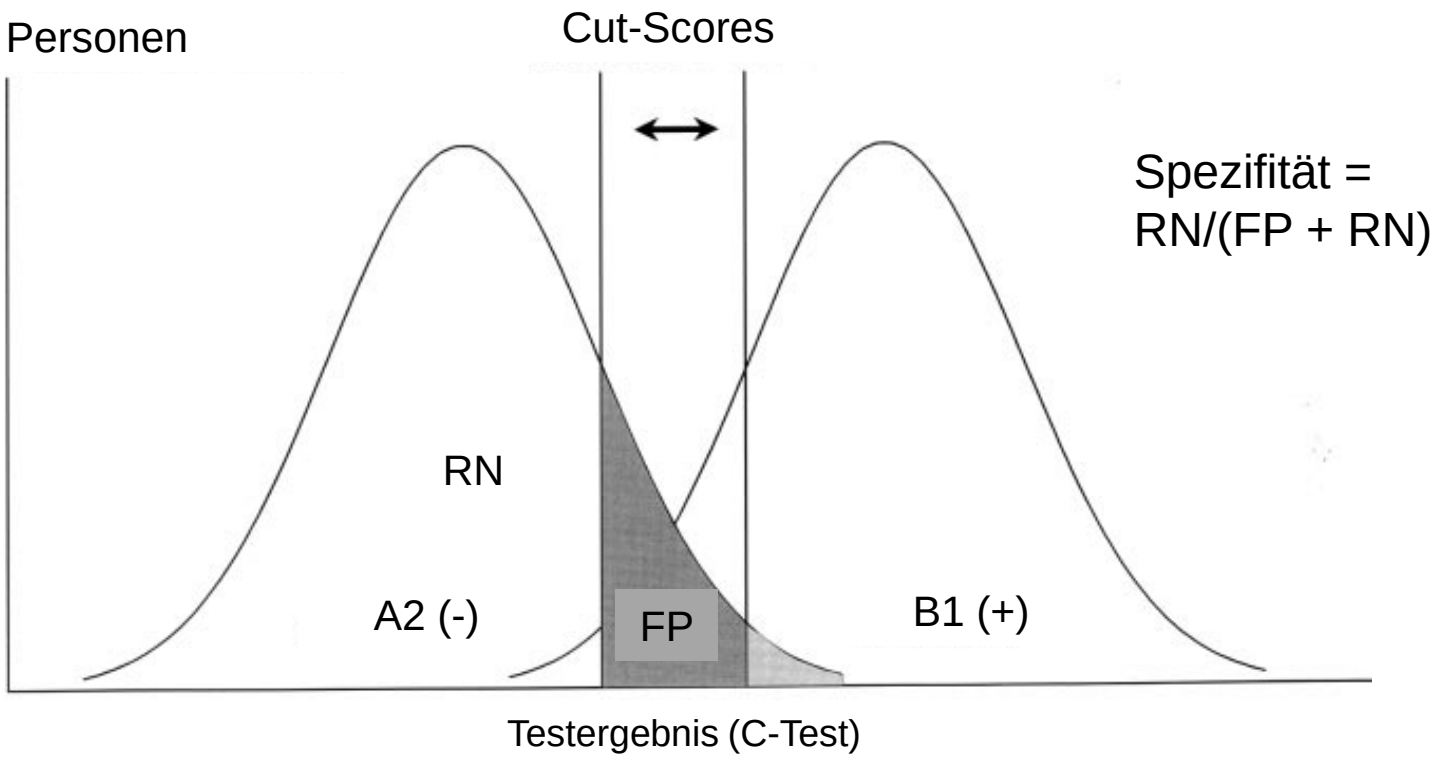


g.a.s.t.

onSET
online-Spracheinstufungstest

Prototypgruppenmethode

ROC-Analyse: Spezifität (Quote korr. Ablehnungen)

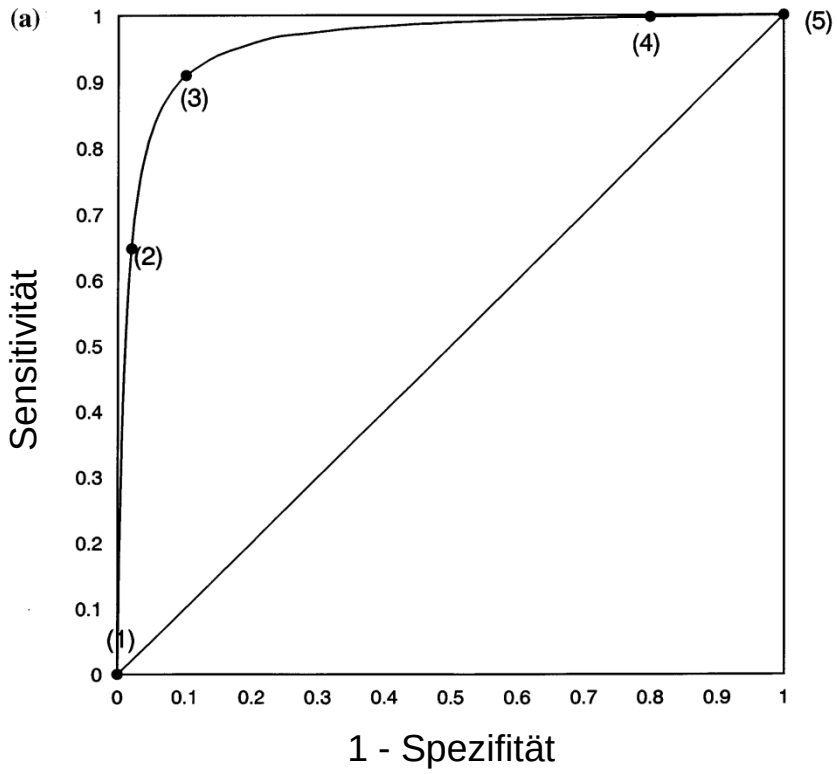


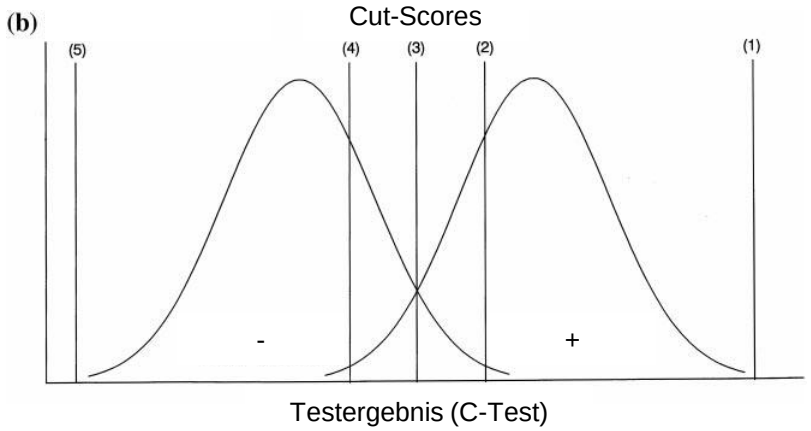
g.a.s.t.

onSET
online-Spracheinstufungstest

Prototypgruppenmethode

ROC-Analyse: ROC-Kurve



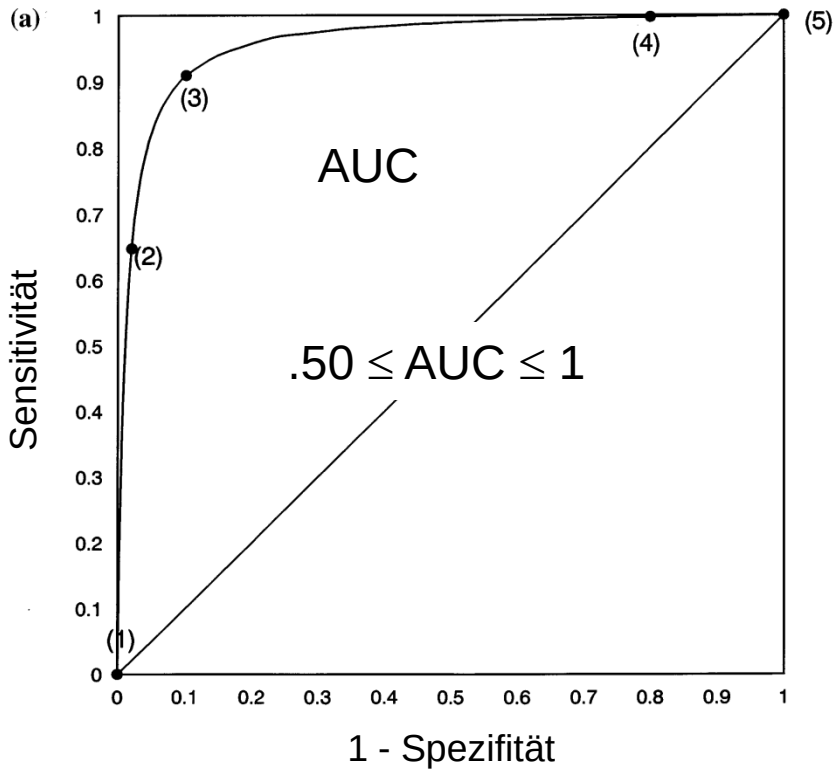


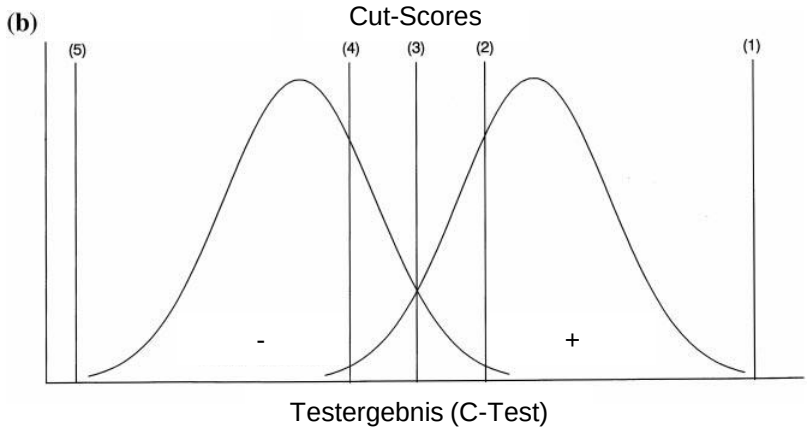
g.a.s.t.

onSET
online-Spracheinstufungstest

Prototypgruppenmethode

ROC-Analyse: AUC (Area Under the Curve)



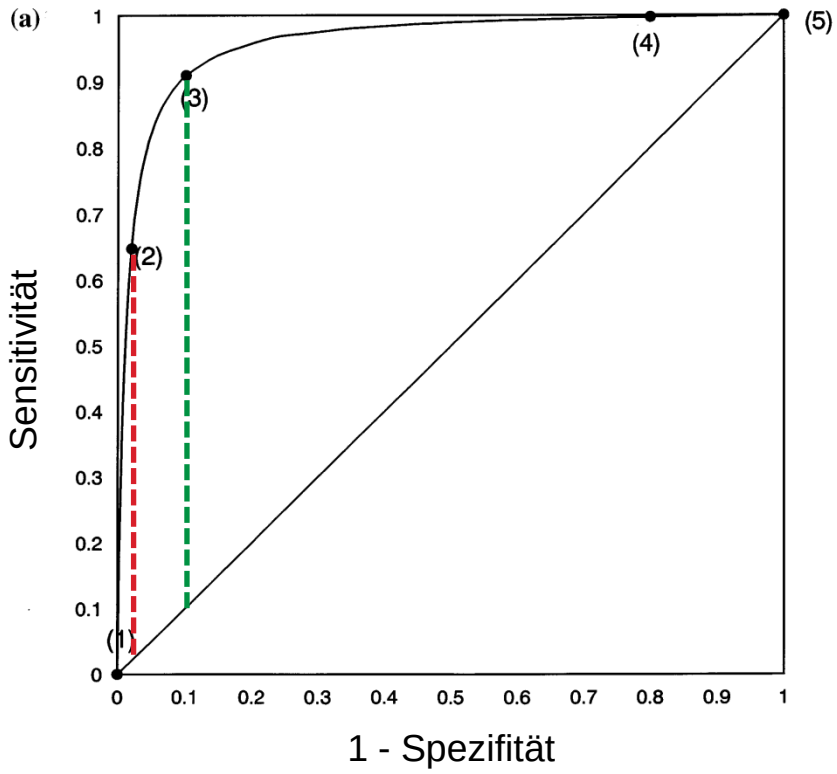


g.a.s.t.

onSET
online-Spracheinstufungstest

Prototypgruppenmethode

ROC-Analyse: Youden-Index *J* (Youden, 1950)



***J* = Sensitivität + Spezifität – 1**

Welcher Cut-Score liefert den höchsten Wert des Youden-Index J?

g.a.s.t.

onSET
online-Spracheinstufungstest

Cut-Scores für onSET-English

Datenbasis

- Weltweite Erprobung neuer (englischer) Texte (*N* = 3.310)
- 20 Sets von je 6 oder 10 Texten
- Ankertestplan (pro Set 2 Ankertexte)
- Benennung von 470 Lernerprototypen (14,2 %)

g.a.s.t.

onSET
online-Spracheinstufungstest

Cut-Scores für onSET-English

ROC-Analyse

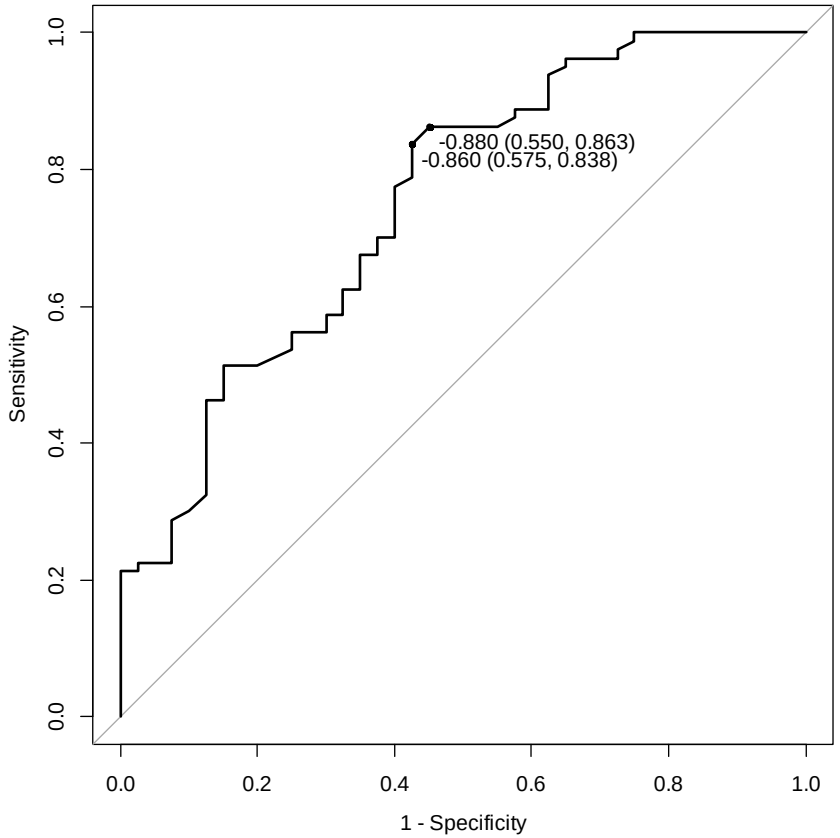
- Niveauvergleiche (gesucht: 4 Cut-Scores)
  - A1 vs. A2 (n = 120)
  - A2 vs. B1 (n = 183)
  - B1 vs. B2 (n = 241)
  - B2 vs. C1 (n = 201)
- Korrelation Prototypen (A1 bis C1) mit Fähigkeitsschätzungen (nach Rasch-Modell): $r(470) = .66$ ($p < .001$)
- Analyse mit R-Paket pROC (Robin et al., 2011)

g.a.s.t.

onSET online-Spracheinstufungstest

Cut-Scores für onSET-English

ROC-Analyse



A1 vs. A2

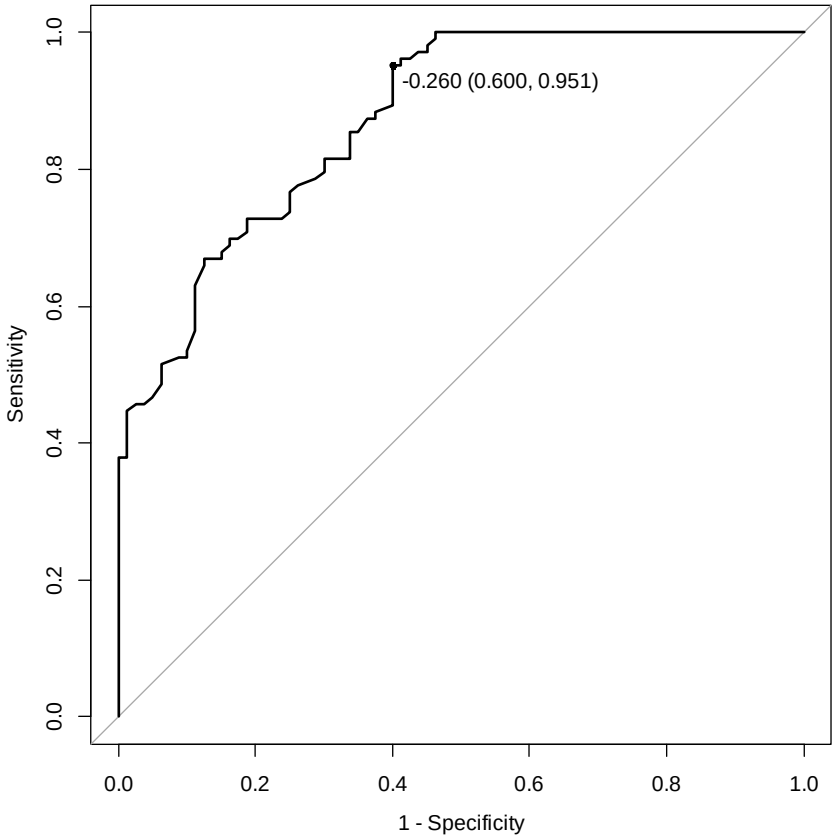
J-Index = 0.41
Cut-Score c_J = -0.86
AUC = 0.75
CI (AUC): 0.66 / 0.84
PCTR = 0.75

g.a.s.t.

onSET
online-Spracheinstufungstest

Cut-Scores für onSET-English

ROC-Analyse



A2 vs. B1

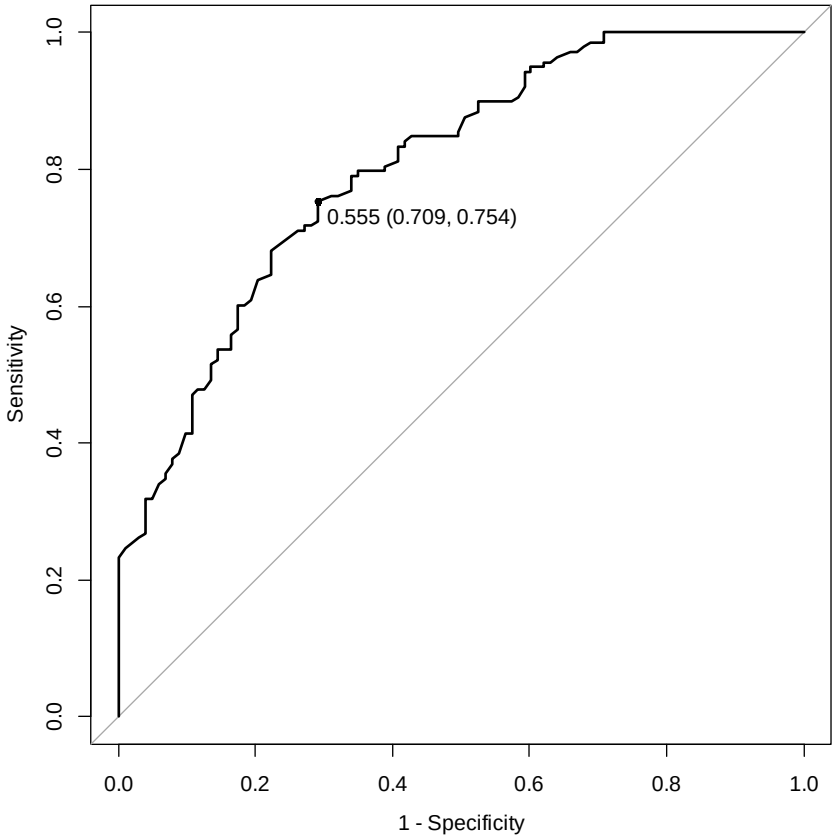
J-Index = 0.55
Cut-Score c_J = -0.26
AUC = 0.87
CI (AUC): 0.82 / 0.92
PCTR = 0.80

g.a.s.t.

onSET
online-Spracheinstufungstest

Cut-Scores für onSET-English

ROC-Analyse



B1 vs. B2

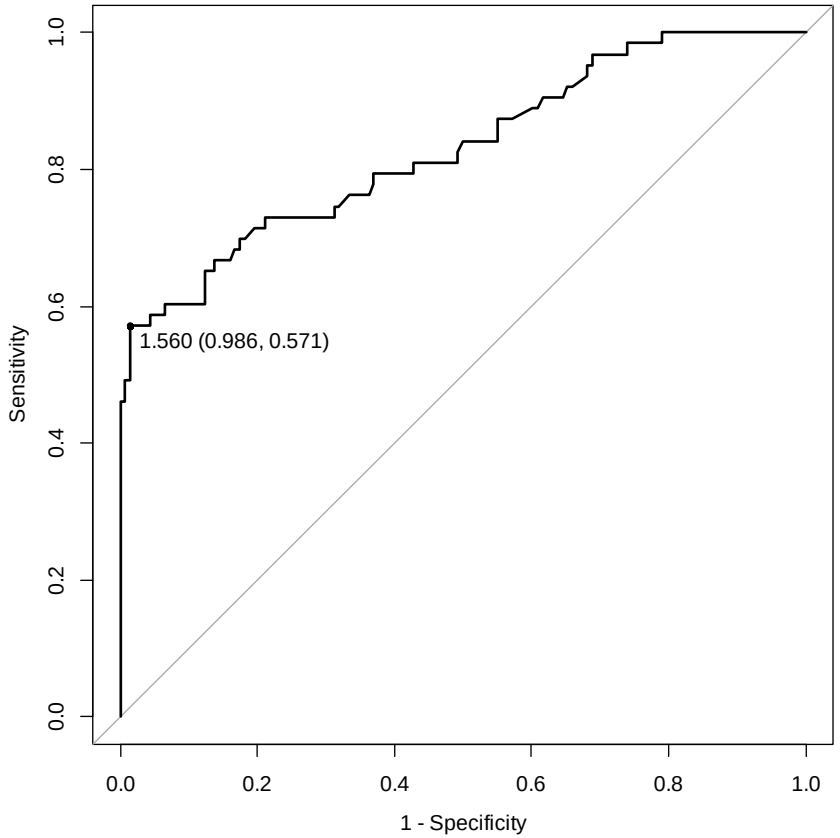
J-Index = 0.46
Cut-Score c_J = 0.55
AUC = 0.80
CI (AUC): 0.75 / 0.86
PCTR = 0.73

g.a.s.t.

onSET
online-Spracheinstufungstest

Cut-Scores für onSET-English

ROC-Analyse



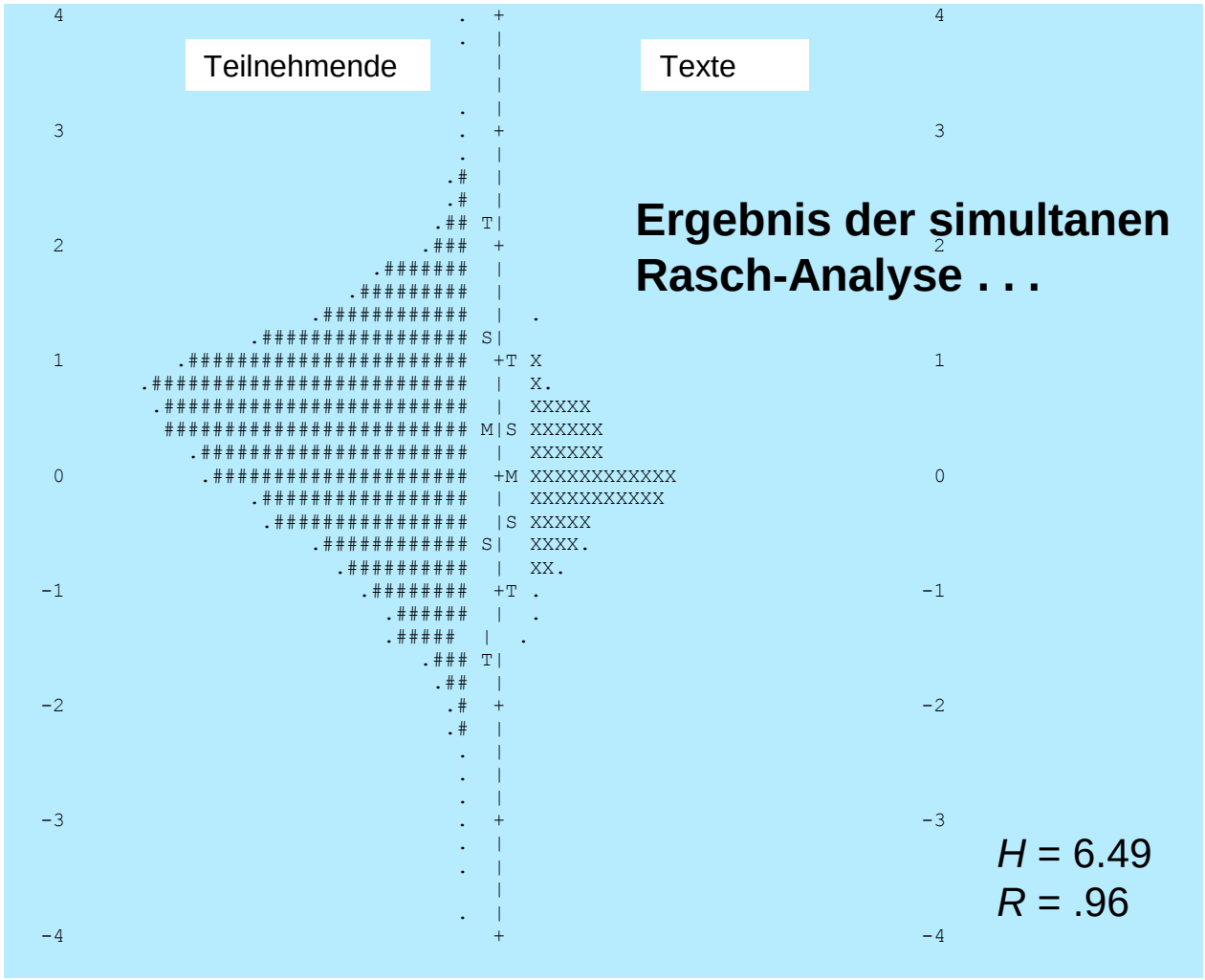
B2 vs. C1

J-Index = 0.56
Cut-Score c_J = 1.56
AUC: 0.83
CI (AUC): 0.76 / 0.89
PCTR = 0.86

g.a.s.t.

onSET
online-Spracheinstufungstest

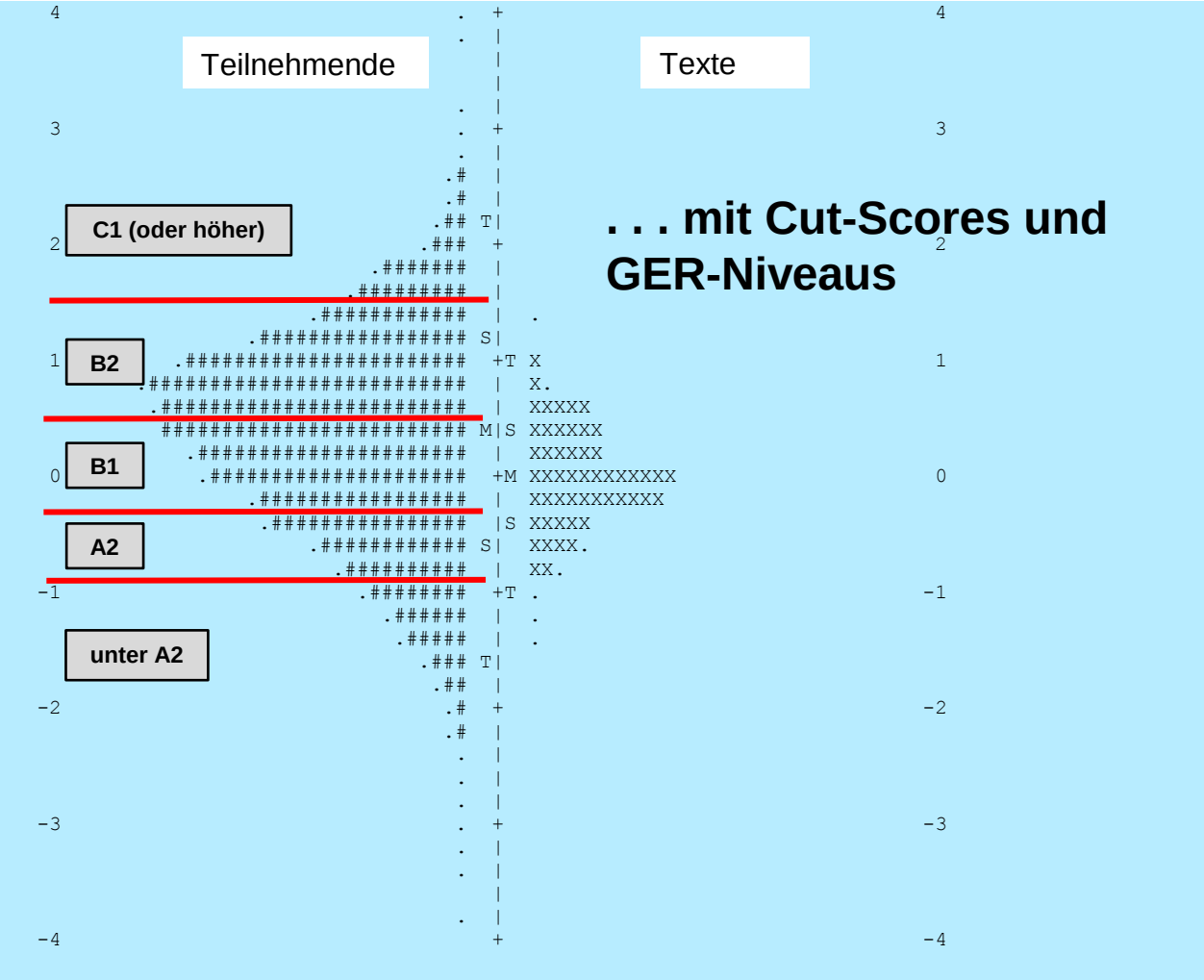
Cut-Scores für onSET-English



g.a.s.t.

onSET online-Spracheinstufungstest

Cut-Scores für onSET-English



g.a.s.t.

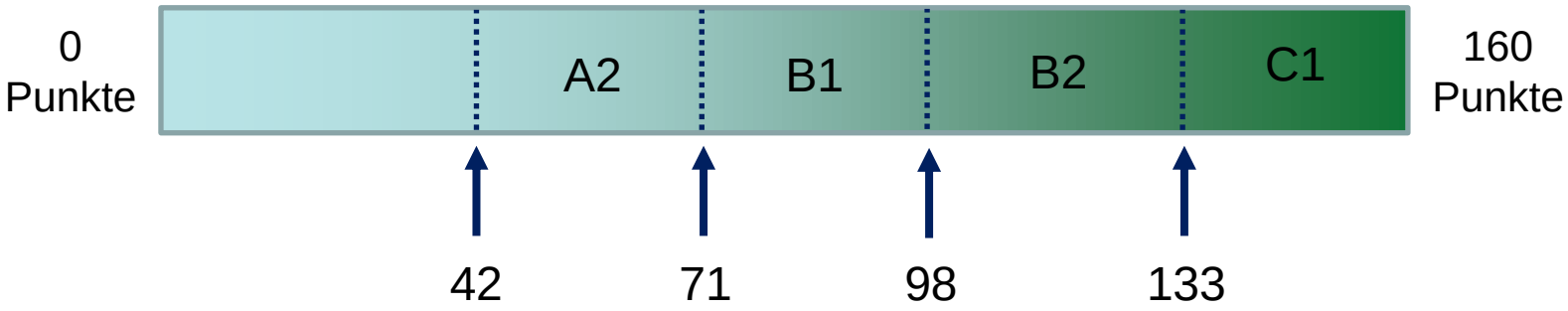
onSET
online-Spracheinstufungstest

Wo liegen die Grenzen zwischen benachbarten Niveaus beim onSET-English?

g.a.s.t.

onSET
online-Spracheinstufungstest

Hier!



g.a.s.t.

onSET
online-Spracheinstufungstest

Zusammenfassung

- Die PGM erlaubt die zuverlässige Bestimmung von GER-Niveaus (nicht nur bei C-Tests)
- In Kombination mit einer ROC-Analyse liefert die PGM trennscharfe Cut-Scores (hohe Genauigkeit der Zuordnung zu GER-Niveaus)
- Dies erweitert die Einsatzmöglichkeiten von C-Tests, insbesondere des onSET
  - onSET für Flüchtlinge, seit April 2016 (refugees.onset.de)
  - Weitere Sprachversionen in Planung (onSET-Français, onSET-Español, onSET-Italiano etc.)

g.a.s.t.

onSET
online-Spracheinstufungstest

Danke.

E-Mail: thomas.eckes@testdaf.de

www.testdaf.de

www.onset.de

g.a.s.t.

onSET
online-Spracheinstufungstest