

g.a.s.t.

# Integrierte Prüfungsleistungen mit dem GER beurteilen

Was die Kann-Beschreibungen  
nicht können

Daniela Marks &  
Sonja Zimmermann  
Bremer Symposion, 25.2.2017



# Inhalte

- Überblick Studie zur Skalenentwicklung
- Integriert Prüfen und Testen
- „Zusammenfassen“ als integrierte Aufgabe
- Beurteilen mit dem GER – (Wie) geht das?
- Erste Ergebnisse aus der Pilotierung der Skalen
- Diskussion & Ausblick



# Überblick Studie Skalenentwicklung

# Überblick Studie zur Skalenentwicklung

- Entwicklung neuer integrierter Aufgabenformate zum Schreiben und Sprechen
- Entwicklung von Beurteilungsskalen für diese Aufgaben ausgehend vom GER
- Pilotierung der Skalen im Rahmen der Erprobung der Aufgaben
- erste qualitative und quantitative Auswertungen der Beurteilungen (*work in progress*)



# Integriert Prüfen und Testen

# Ausgangslage

*„promising  
alternative“*

(Guo, Crossley & McNamara, 2013)

*„recent  
trend“*

(Weigle & Parker, 2012)

*„topic of  
heated  
debates“*

(Yu, 2013)

integrierte  
Prüfungsaufgaben

*„increasingly  
used“*

(Plakans & Gebriel, 2014)

*„already common  
practice“*

(Eckes, Müller-Karabil & Zimmermann, 2016)

*„much under-  
researched“*

(Li, 2014)

# Definition integrierte Prüfungsaufgaben

## in Abgrenzung zu isolierten Prüfungsaufgaben:

„The term ‚integrated‘ has been applied fairly broadly to tasks that require more than one skill for completion.“

Plakans, 2013

„In contrast, integrated tasks require the test-taker to integrate multiple language skills **in a substantial way** to complete a [...] task“

Lee, 2006

# Definition integrierte Prüfungsaufgaben

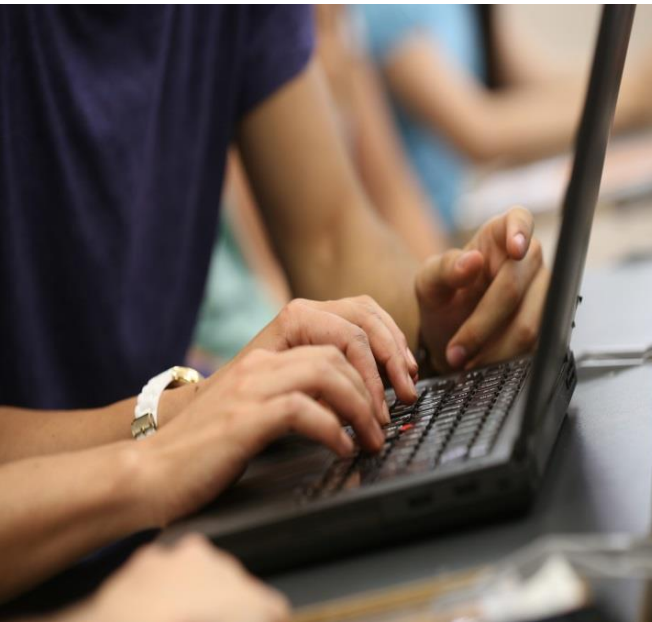
**(Schreib-)Aufgaben sind dann integriert, wenn**

- (1) der Input einen gewissen Umfang an Sprachmaterial beinhaltet,
- (2) sich der Output inhaltlich auf den Input bezieht und
- (3) die Sprache des Inputs umgewandelt werden muss.

„the stimulus materials need **to provide sufficient language** (either in written or audio format) to allow writers [and speakers, sz&dm] **to produce sufficient text** to be rated by assessors“

Knoch & Sitajalabhorn, 2013



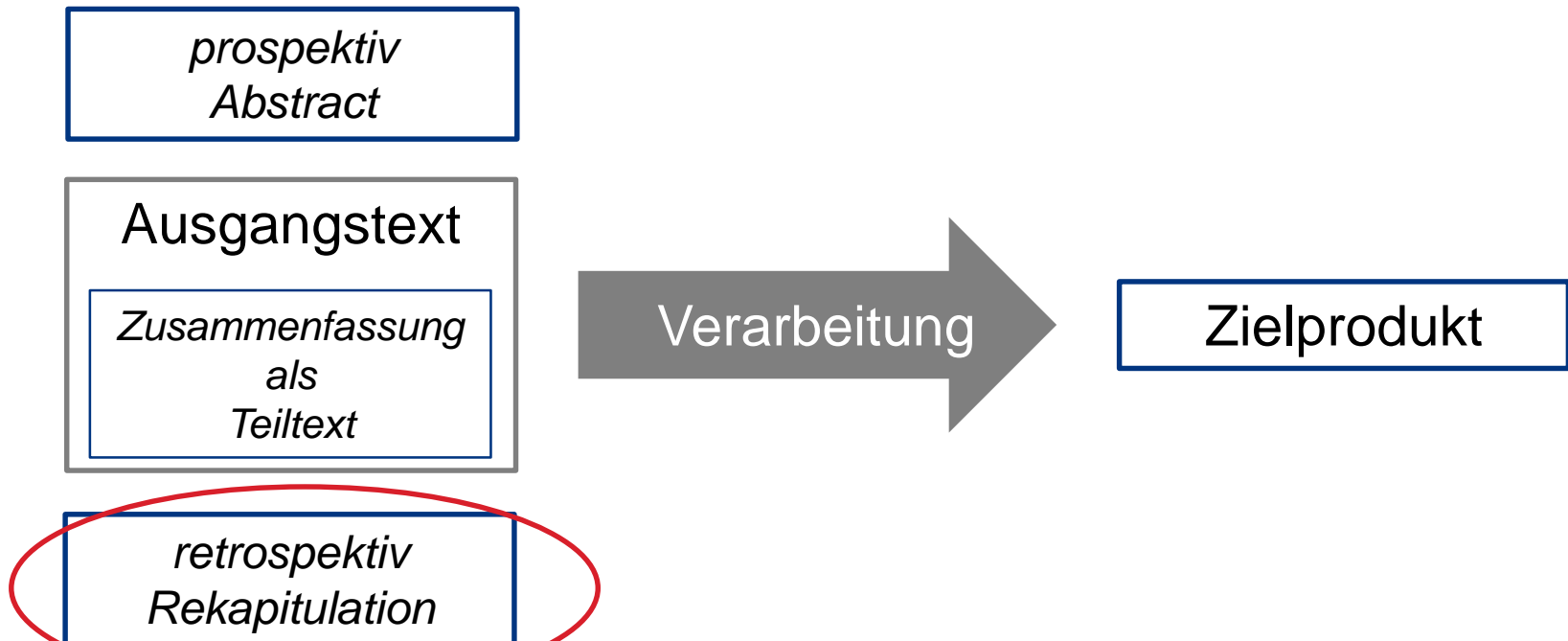


# „Zusammenfassen“ als integrierte Aufgabe

# Schreib-/Sprechhandlung „Zusammenfassen“

## Zusammenfassung im akademischen Kontext

- als Textsorte meist bezogen auf schriftliche Texte
- „Hilfstextart“ vs. eigenständige Textsorte



Ehlich, 2003; Stezano Cotelo, 2008



# Beurteilen mit dem GER – (Wie) geht das?

# Entwicklung von Beurteilungsskalen

## GER-Skalen als Ausgangspunkt

- GER-Skalen sind eher **benutzerorientiert** und enthalten daher sehr allgemeine, wenig spezifische und abstrakte Beschreibungen von Sprachkompetenz
  - **Beurteilungsskalen** dagegen erfordern eine sehr genaue Beschreibung relevanter Aspekte der Schreib-/Sprechkompetenz bezogen auf den jeweiligen Kontext
- Anpassung notwendig (Galaczi et al., 2011; Harsch & Martin, 2012; Deygers & Van Gorp, 2015), vor allem mit Blick auf integrierte Aufgaben, die besondere kognitive Verarbeitungsprozesse verlangen

# Zusammenfassung / Quellenmaterial verarbeiten im GER



Texte verarbeiten (GER S. 98)	
C2	Kann Informationen aus verschiedenen Quellen <b>zusammenfassen</b> und die Argumente und berichteten Sachverhalte so wiedergeben, dass insgesamt eine kohärente Darstellung entsteht.
C1	Kann lange, anspruchsvolle Texte <b>zusammenfassen</b> .
B2	Kann ein breites Spektrum von Sachtexten und fiktiven Texten <b>zusammenfassen</b> und dabei die Hauptthemen und unterschiedliche Standpunkte kommentieren und diskutieren. Kann Auszüge aus Nachrichten, Interviews oder Reportagen, welche Stellungnahmen, Erörterungen und Diskussionen enthalten, <b>zusammenfassen</b> . Kann die Handlung und die Abfolge der Ereignisse in einem Film oder Theaterstück <b>zusammenfassen</b> .
B1	Kann kurze Informationen aus mehreren Quellen zusammenführen und für jemand anderen <b>zusammenfassen</b> .
	Kann kurze Textpassagen auf einfache Weise <b>zusammenfassen</b> , indem er /sie dabei den Wortlaut und die Anordnung des Originals benutzt.

# Was kennzeichnet eine Zusammenfassung?

## ZUSAMMENFASSUNG:

- kürzer als Originaltext / Beschränkung auf wesentliche Punkte / Unterscheiden von wichtigen und unwichtigen Details
- nachvollziehbar strukturiert, an Inputtext in der logischen Struktur angelehnt
- Nur Aussagen aus Text (keine eigenen Ergänzungen)
- Darstellung der Textstruktur auf Metaebene

- Struktur
- wichtigste Informationen selektieren
- wenig Details (Beispiele)

# Welche Kompetenzen sind erforderlich?

- Grafik-
- Textverständnis
  - Umformulierungen, <sup>eigenständig</sup> formulieren
  - Strukturierende spr. Mittel (Konnektoren)
  - Register (für wen ist Zsf.?)
- allgemein: Distanz zum Text

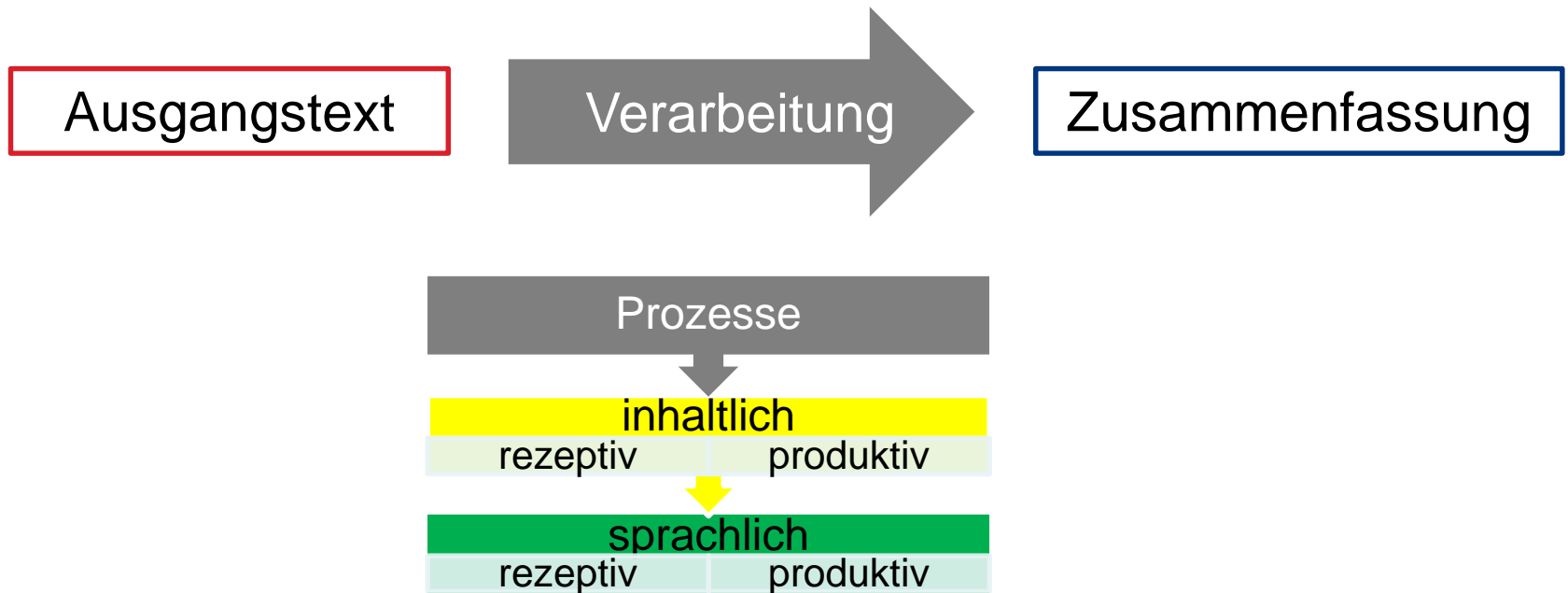
## SPRACHLICHE KOMPETENZEN:

- Paraphrasieren, Text(teile) verdichten (~~etwas~~)
- breites Spektrum an Wortschatz und Strukturen  
→ z.B. Kohäsionsstiftende Mittel
- Input angemessen verarbeiten (Hv, Lv, kognitiv)
- Struktur des Inputtexts analysieren

⇒ Inhalte verdichtet darstellen (z.B. durch Synonyme, synt. Umformungen ...)

# Beurteilung Zusammenfassung

Wie kann man die geforderte inhaltliche Verarbeitung und die sprachliche Transformation des Inputs in den Beurteilungskriterien spiegeln?





# Beurteilung Zusammenfassung

**Schreiben:** Abbildung der Verarbeitungsprozesse in den Beurteilungskriterien

Kriterium	Aspekte (Rezeption & Produktion)
<ul style="list-style-type: none"><li>▪ <b>Verständnis des Inputs</b></li><li>▪ <b>Zusammenfassung</b></li></ul>	<ul style="list-style-type: none"><li>▪ relevante Informationen aus Ausgangstext enthalten</li><li>▪ korrekte Wiedergabe der Informationen</li><li>▪ Struktur</li></ul>
<ul style="list-style-type: none"><li>▪ <b>Sprachliche Mittel</b></li></ul>	<ul style="list-style-type: none"><li>▪ Spektrum sprachlicher Mittel zur eigenständigen Formulierung</li><li>▪ präzise Verwendung relevanter Begriffe aus dem Ausgangstext</li></ul>

# Beurteilung Zusammenfassung

**Sprechen:** Abbildung der Verarbeitungsprozesse in den Beurteilungskriterien

Kriterium	Aspekte (Rezeption & Produktion)
<ul style="list-style-type: none"><li>▪ <b>Inhaltliche Zusammenfassung</b></li><li>▪ <b>Struktur</b></li></ul>	<ul style="list-style-type: none"><li>▪ wesentliche Informationen aus Ausgangstext enthalten</li><li>▪ korrekte Zusammenfassung der Informationen</li><li>▪ Klare und logische Struktur</li></ul>
<ul style="list-style-type: none"><li>▪ <b>Wortschatz</b></li></ul>	<ul style="list-style-type: none"><li>▪ Präzision des Wortschatzes (auch aus dem Inputtext), Umgang mit Lücken im Wortschatz</li><li>▪ Wortschatzfehler</li></ul>



# Ergebnisse aus der Pilotierung

# Teilnehmende – Demografische Daten

	<b>Schreiben</b>	<b>Sprechen</b>
<b>N</b>	143	110
<b>Alter (Jahre)</b>	<i>M</i> = 25,4	<i>M</i> = 25,1
<b>Geschlecht (Prozent)</b>	53,0% w; 47,0% m	57,3% w ; 42,7% m
<b>Top Herkunftsländer</b>	Syrien 20 China 18 Russland 13 Ukraine 6 Kolumbien 5	China 21 Syrien 11 Ägypten 4 Ukraine 3 Südkorea 3

# Teilnehmende – Einstufungen im onSET

	Schreiben	Sprechen	Kernbereich
<b>C1 oder höher</b>	9,0 %	3,7 %	135-140 Punkte
<b>B2.2</b>	18,7%	14,6%	106-111 Punkte
<b>B2</b>	45,6 %	50 %	
<b>B2.1</b>	26,9%	35,4 %	
<b>B1</b>	38,7 %	45,1%	74-79 Punkte
<b>A2 oder niedriger</b>	6,7 %	1,2 %	46-51 Punkte

# BeurteilerInnen

- 10 erfahrene *rater*: ExpertInnen im Bereich Skalenentwicklung, HochschuldozentInnen, mit Zielgruppe vertraute BeurteilerInnen

Verteilung der Leistungen auf die BeurteilerInnen:

- Leistungen jeweils von zwei BeurteilerInnen beurteilt
- jede/r Beurteiler/in hat mit vier anderen Überschneidungen

# Ergebnisse – GER-Stufen nach Facettenanalyse

	<b>schriftliche Zusammen- fassung.</b>	<b>mdl. Zusammen- fassung (Lesetext)</b>	<b>mdl. Zusammen- fassung (Hörtext)</b>
C1	0,0 %	4,5 %	2,7 %
B2.2	10,5 %	11,8 %	12,7 %
B2.1	35,0 %	32,7 %	42,7 %
unter B2	54,5 %	50,9 %	41,8 %

# Ergebnisse

## Statistische Analysen

- Konsistenz der Beurteiler zufriedenstellend
- Strenge und Milde der Beurteiler nach Facettenanalyse im akzeptablen Bereich
- Kriterien, die sich auf **sprachliche Aspekte** beziehen, funktionieren weitgehend zufriedenstellend
- auffällige Infit-/Outfit-Werte insbesondere bei den **inhaltlichen Kriterien**



# Ergebnisse

## Schreiben: Rückmeldungen der BeurteilerInnen

- „manchmal schwierig, die beiden inhaltlichen Kriterien auseinanderzuhalten“
- „Wie bewerten, wenn der Text gut strukturiert ist, aber keine gelungene Zusammenfassung?“
- „Überlappung der Kriterien *Zusammenfassung* und *Kohärenz & Kohäsion*“
- „Unterscheidung von *Zusammenfassung* und *Kohärenz & Kohäsion* fällt schwer, in beiden geht es oft um Strukturierung“

# Ergebnisse

## Sprechen: Rückmeldungen der BeurteilerInnen

- „Wie bewerten, wenn nur die Hauptaussage des Textes genannt wird und ansonsten eigenes Wissen wiedergegeben wird?“
- „Wie mit der Übernahme syntaktischer Strukturen aus dem Inputtext umgehen?“
- Aufgabe mit gehörtem Input: „quantitative Aspekte in beiden inhaltlichen Kriterien“

Kriterium	Aspekte (Rezeption & Produktion)
<ul style="list-style-type: none"><li>▪ <b>Inhaltliche Zusammenfassung</b></li></ul>	<ul style="list-style-type: none"><li>▪ wesentliche Informationen aus Ausgangstext enthalten</li><li>▪ korrekte Zusammenfassung der Informationen</li></ul>
<ul style="list-style-type: none"><li>▪ <b>Vollständigkeit</b></li></ul>	<ul style="list-style-type: none"><li>▪ Berücksichtigung der vorgegebenen Punkte</li></ul>



# Diskussion & Ausblick

# Diskussion

## Einschränkungen

- relativ schwaches Sample
- unbekanntes Aufgabenformat, keine Vorbereitung
- keine Schulung der *rater* in der Anwendung der Skalen
- keine Standardisierung der Beurteilung durch Kalibrierung

# Ausblick

## **Nächste Schritte**

- weitere (statistische) Auswertungen
- linguistische Analysen der Teilnehmerleistungen
- Überarbeitung der Skalen und Deskriptoren
- weitere Erprobung mit neuen Aufgabeninhalten, größerem Sample (v.a. auf B2.2/C1-Niveau)
- Studien zu Einflussfaktoren

# DANKE!

[Sonja.Zimmermann@testdaf.de](mailto:Sonja.Zimmermann@testdaf.de)

[Daniela.Marks@testdaf.de](mailto:Daniela.Marks@testdaf.de)

**g.a.s.t.**

# Literatur

- Deygers, B. & Van Gorp, K. (2015). Determining the scoring validity of a co-constructed CEFR-based rating scale. *Language Testing*, 32(4), 521–541.
- Eckes, T., Müller-Karabil, A. & Zimmermann, S. (2016). Assessing writing. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 147–164). Boston, MA: De Gruyter.
- Ehlich, K. (2003). Universitäre Textarten, universitäre Struktur. In K. Ehlich & A. Steets (Hg.), *Wissenschaftlich schreiben – lehren und lernen* (S. 13–28). Berlin/New York: de Gruyter.
- Galaczi, E.; French, A.; Hubbard, C. & Green, A. (2011): Developing assessment scales for large-scale speaking tests: a multiple-method approach. *Assessment in Education: Principles, Policy & Practice*, 18 (3), 217–237.
- Gemeinsamer europäischer Referenzrahmen für Sprachen (2001).
- Guo, L.; Crossley, S. & McNamara, D. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing* 18, 218–238.
- Harsch, C. & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing* 17, 228–250.
- Knoch, U. & Sitjalabhorn, W. (2013). A closer look at integrated writing tasks. Towards a more focused definition for assessment purposes. *Assessing Writing* 18 (4), 300–308.
- Lee, Y. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing* 23 (2), 131–166.
- Li, J. (2014). The role of reading and writing in summarization as an integrated task. *Language Testing in Asia* 4 (3).
- Plakans, L. (2013). Assessment of integrated skills. In Chapelle, C. (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 205-212). Malden, MA: Wiley-Blackwell.
- Plakans, L. & Gebril, A. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing*, 21 (2), 56–73.
- Stezano Coteló, K. (2008). *Verarbeitung wissenschaftlichen Wissens in Seminararbeiten ausländischer Studierender. Eine empirische Sprachanalyse*. München: Iudicum.
- Weigle, S. C.; Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing* 21 (2), 118–133.
- Yu, G. (2013): From integrative to integrated language assessment. Are we there yet? *Language Assessment Quarterly* 10 (1), 110–114.