



UNIVERSITÄT
KOBLENZ · LANDAU



**MERCATOR
INSTITUT**
für Sprachförderung
und Deutsch
als Zweitsprache

WORKSHOP 4

Schreibkompetenzen messen

Referent*innen:

Dr. Ann-Kathrin Hennes, Universität zu Köln

Junior-Prof. Dr. Markus Linnemann, Universität Koblenz-Landau

Vorstellung

Ann-Kathrin Hennes

- Dissertation (2020) *Schreibprodukte bewerten – Die Rolle der Expertise bei der Bewertung der Textproduktionskompetenz*
- Projekt: Schreibtest 4-9
- Projekt: Überarbeitungsstrategien während des Schreibprozesses & der Zusammenhang zu Textqualität (unter Berücksichtigung des AG)

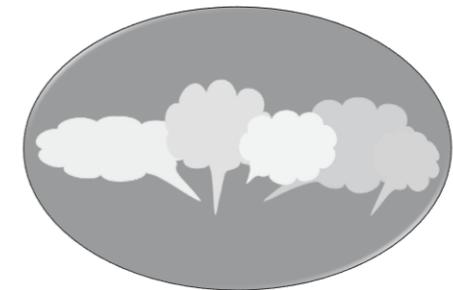
Markus Linnemann

- Dissertation (2014) *Kognitive Prozesse der Adressatenantizipation beim Schreiben*
- Projekt: Schreibtest 4-9
- Projekt: Lese- und Schreibflüssigkeit - Konzeption, Diagnostik, Förderung (FluLeS)
- U.v.m.

Vorstellung

Sie sind dran 😊: Eine erste Breakout Session

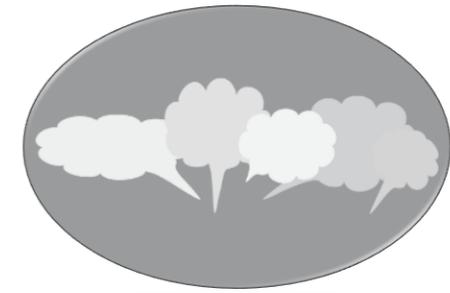
- Welche Texte bewerten Sie?
- Welche Aspekte berücksichtigen Sie bei der Textbewertung?
- Dauer der Breakout Session: 10 Minuten
- Anschließend kurzer Austausch



Vorstellung

Austausch

- Welche Texte bewerten Sie?
- Welche Aspekte berücksichtigen Sie bei der Textbewertung?



Relevanz der Schreibkompetenz

- **Schlüsselkompetenz für gesellschaftliche Teilhabe**
(Becker-Mrotzek, 2014)
- **Entscheidende Voraussetzung für schulischen, akademischen und beruflichen Erfolg**
(Crossley & McNamara, 2016; MacArthur, Graham & Fitzgerald, 2016)
- **Sonderpädagogik: $\frac{1}{4}$ bis $\frac{3}{4}$ der Personen mit Lernstörungen in Ausbildung müssen in ihrem Beruf zumindest kurze Texte verfassen**
(Bach, Schmidt, Schabmann & van Koll, 2016)

Vermittlung der Schreibkompetenz

Es braucht:

- Schreibunterricht mit Zeit und Raum für Übung
(Langer, 2017)
- Unterricht entlang der Stärken & Schwächen der Schüler*innen
(Beck, Llosa, Black & Anderson, 2018)
- Symptomspezifische Förderung
(Graham & Perin, 2007)
- formative Evaluation erfolgter Maßnahmen
(Hattie, 2009)

Kernlehrplan NRW (2012):

- Lernstand und Lernentwicklung der Schülerinnen und Schüler beobachten
- Wirkung von Unterrichts- und Fördermaßnahmen evaluieren
- Differenzierung des Unterrichts anhand der individuellen Lernausgangslagen
- Zieldifferentes Unterrichten

Messung der Schreibkompetenz

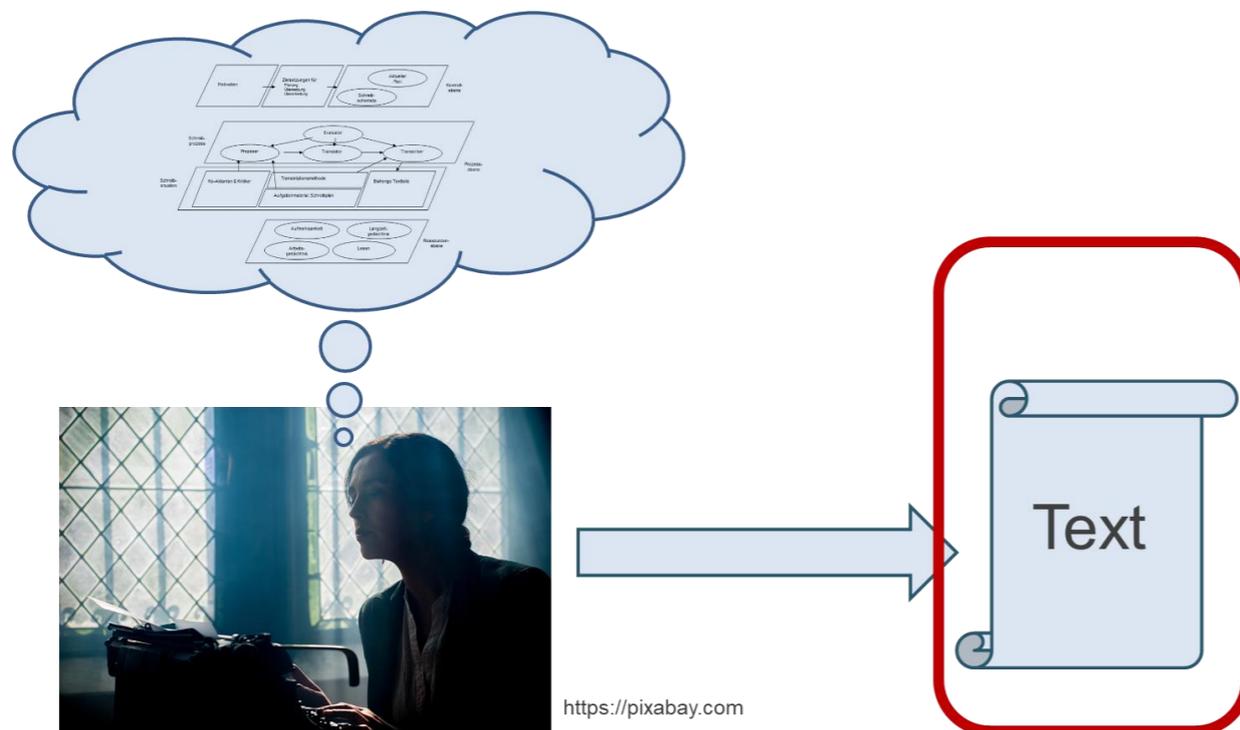
Wanted: objektive, reliable und valide Bewertungskriterien

→ akkurate & differenzierte Diagnostik

Messung der Schreibkompetenz

Vorgehen in Schule & Forschung

- Schreibkompetenz wird über Messung der Textqualität bestimmt (Feenstra, 2014; Philipp, 2015)
- Hierzu Bewertung (längerer) Texte
 - global → als Ganzes (Eckes, 2008; Skar & Jølle, 2017)
 - analytisch → mithilfe von Kriterien (Eckes, 2008; Weigel, 2002)



<https://pixabay.com>

Messung der Schreibkompetenz

Probleme bei der schulischen Textbewertung

- Fehlende Übereinstimmung zwischen Lehrkräften bei der Bewertung des gleichen (Schüler*innen-)Textes (Birkel & Birkel, 2002; Cooksey, Freebody & Wyatt-Smith, 2007; Skar & Jølle, 2017)
- Beeinträchtigung der Objektivität und Reliabilität

Messung der Schreibkompetenz

Mögliche Gründe für fehlende Übereinstimmung

- **Ambiguität vorhandener „Textbewertungskriterien“**
(Knoch, 2011; Lumley, 2002; Todd, Thienpermpool & Keyuravong, 2004)
- **Unterschiede in der Interpretation oder Gewichtung der Kriterien** (Cooksey, Freebody & Wyatt-Smith, 2007; Eckes, 2008)
- **Nutzung sozialer Norm**
(Ingenkamp & Lissman, 2008; Cooksey, Freebody & Wyatt-Smith, 2007)
- **Nutzung theoretisch weniger relevanter Kriterien**
(Rezaei & Lovorn, 2010; Vögelin et al., 2018; 2019)

Messung der Schreibkompetenz

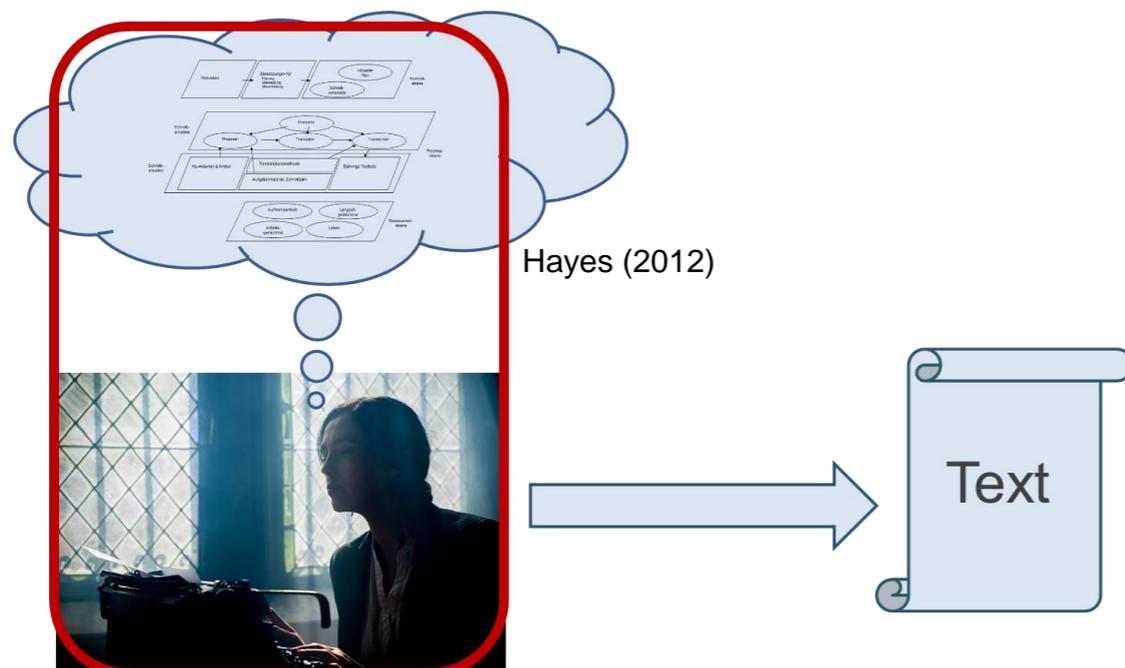
Textbewertung in der Forschung

- Entwicklung von Bewertungskriterien (analytische Skalen, Benchmarktexte)
- (aufwendiges) Training der Bewerter*innen in der Anwendung dieser Skalen (e.g. Brown et al., 2004; Baird et al., 2004; Eckes, 2008; Lim, 2011; van Gasse et al., 2019)
- Effekte des Einsatzes spezieller Bewertungsskalen & Training der Bewerter*innen sind allerdings begrenzt (Jølle, 2014; Weigle, 1998)
- Auch Schreibforscher*innen stimmen in Ihren Urteilen zur Textqualität nicht vollständig miteinander überein; Konsistenz bei der Textbewertung ist vergleichbar mit der von Lehrkräften (Hennes, 2020)
- Gruppe von Experten der Textbewertung ist sehr klein (Hennes, in Vorbereitung)

Messung der Schreibkompetenz

Implikationen

- Identifikation und Definition manifester Minimalanforderungen (!) für einen guten Text → Entwicklung eindeutiger Textbewertungskriterien
- **Entwicklung spezifischer Schreibaufgaben, die die direkte Messung einzelner Komponenten/Dimensionen der Schreibkompetenz erlauben & psychometrischen Gütekriterien entsprechen**



Zielsetzung

- Entwicklung eines (förder-)diagnostischen Testverfahrens zur Erfassung der Schreibkompetenz für Schüler*innen der 4. bis 9. Jahrgangsstufe.
- Möglichst präzise Erfassung des Kernbereichs von Schreibkompetenz (60 Minuten Testzeit; Gruppentest)
- Orientierung an Kriterien testtheoretischer Güte
- Differenzierte Erfassung des oberen und unteren Leistungsbereichs

Verfahren zur Messung der Schreibkompetenz

- Diagnostische Instrumente zur Bestimmung von Schreibkompetenz 
 - *Test of Written Language* (Hamil & Larsen, 2009)
 - *Kaufmann Test of Educational Achievement* (Kaufman & Kaufman, 2005)

→ Limitation hinsichtlich erfasster Konstrukte
- Diagnostische Instrumente zur Bestimmung von Lesekompetenz 
 - ELFE 1-6 (Lenhard & Schneider, 2006)
 - LGVT 6-12 (Schneider, Schlagmüller & Ennemoser, 2007)
 - HAMLET 3-4 (Lehmann, Peek & Poerschke, 2006)

Interdisziplinäres Team

- Sonderpädagogik & Erziehungswissenschaft
- Linguistik & Sprachdidaktik
Prof. Dr. Jörg Jost,
Prof. Dr. Michael Becker-Mrotzek
- Psychologie
Prof. Dr. Alfred Schabmann
Prof. Dr. Christian Rietz
Vertr.-Prof. Barbara Schmidt
- Sprachtherapie
Anja Obergfell



Methode

Erfassung verschiedener Dimensionen des Konstruktes „Schreibkompetenz“ über verschiedene Aufgabentypen und für das Jahrgangsspektrum geeigneter Textsorten

Textsorten

- Eignung für das Jahrgangsspektrum der 4. bis 9. Jahrgangsstufe: informierende Texte (z.B. Beschreibung, Bericht, Erklärung, Instruktion)

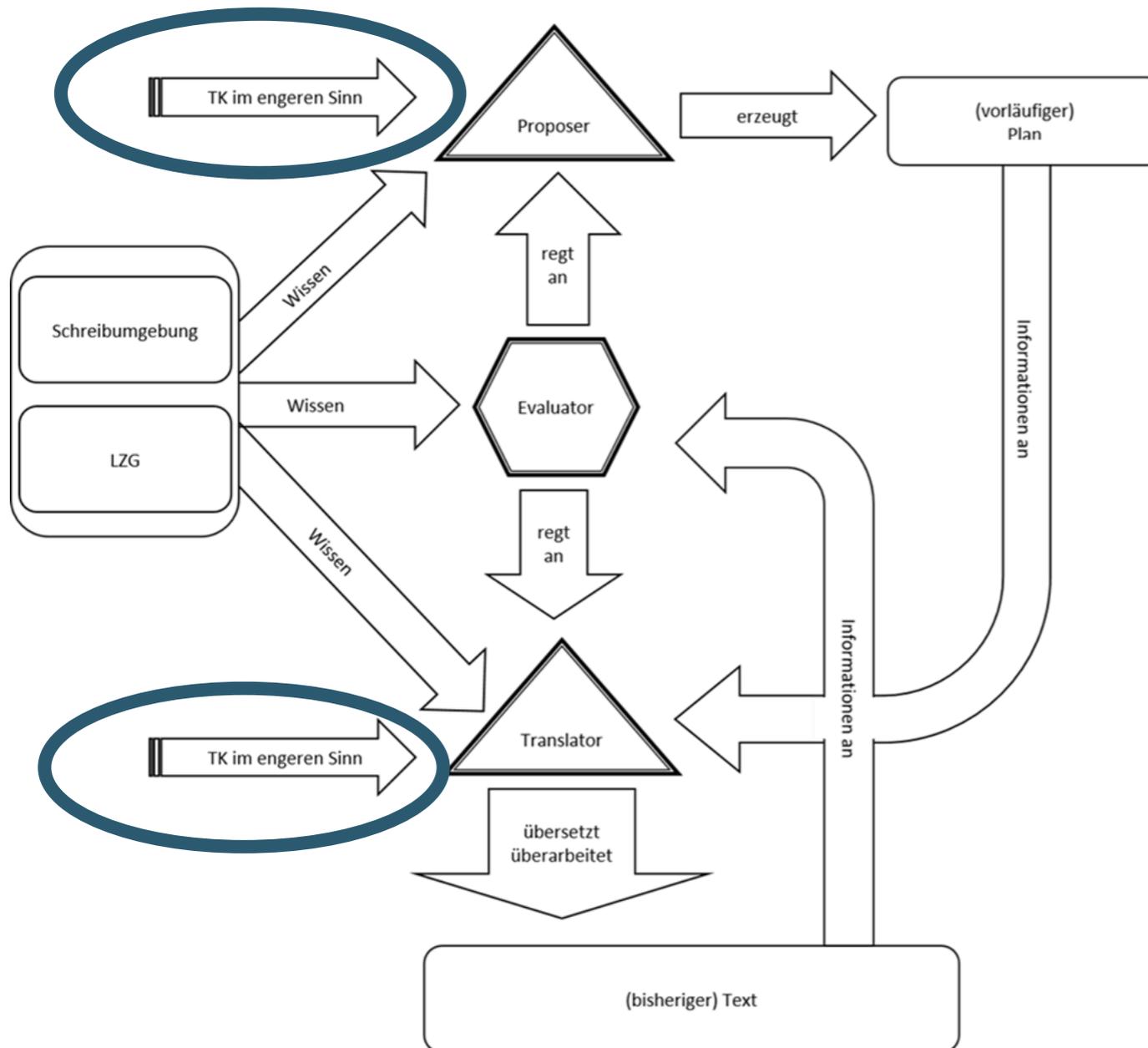
Aufgabentypen

- Aufgaben zur Produktion von ‚Langtexten‘
- Aufgaben geringeren Umfangs mit textproduktiven und -rezeptiven Anforderungen (bspw. Multiple-Choice, Lückentexte) und bezogen auf Teildimensionen des Konstruktes



Konstrukt Schreibkompetenz

Das Konvergenzmodell (Hennes et al., 2018)



Textproduktionskompetenz im engeren Sinne:

- Adressatenorientierung
- Textmusterwissen
- Informationsmanagement
- Kohärenz
- (sprachliche) Kreativität

Konstrukt Schreibkompetenz

Informationsmanagement

- Generierung von Ideen/Informationen
- Angemessene Informationsdichte & Organisation entlang der Grice'schen Maximen (1975):
 - Auswahl relevanter Ideen/Informationen
 - geordnete und widerspruchsfreie Wiedergabe
 - Unterstützung der mentalen Modellbildung & Ermöglichung von Implikaturen (Grice, 1975)
- Informationsweitergabe als kooperativer Prozess

Konstrukt Schreibkompetenz

Informationsmanagement: Subskala Informationsdichte

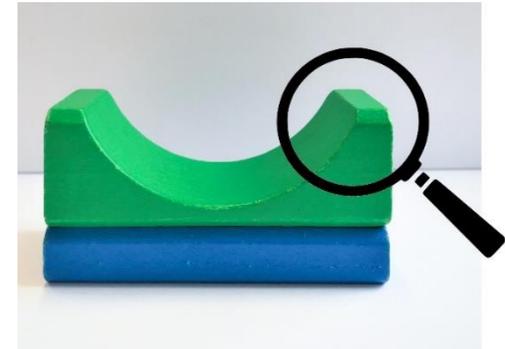
- Vorgabe eines „echten“ Modells (Bausteingebäude), zu dem eine Instruktion geschrieben werden soll.
- Die Instruktion muss dazu geeignet sein, dass Leser*innen anhand des Textes ein mentales Modell konstruieren können, d.h. das Bausteingebäude nachbauen könnten.



Konstrukt Schreibkompetenz

Ver_rückte Gebäude – Subskala adäquate Informationsdichte

(Hennes et al., 2018)



Konstruktionsprinzipien:

- Jedes Gebäude enthält eine Lupe
(kritisches Element, das explizite Informationen erfordert)
 - Bausteine variieren hinsichtlich Farbe & Form nur gemeinsam
 - Steigende Komplexität der Bausteingebäude
 - sachliches Vorwissen a priori im Test vorgegeben
 - Adressatenkreis eindeutig benannt und Wissen explizit beschrieben
- Schau dir das Bild genau an. Baue das Gebäude nach.
 - Überprüfe, ob du alles richtig gebaut hast. Achte besonders auf diese Stelle [Lupe].
 - Ein Freund soll das Gebäude nachbauen. Er hat das Bild nicht. Schreibe ihm eine genaue Anleitung dazu, wie er das Gebäude nachbauen kann.

Konstrukt Schreibkompetenz

Ver_rückte Figuren – Subskala adäquate Informationsdichte

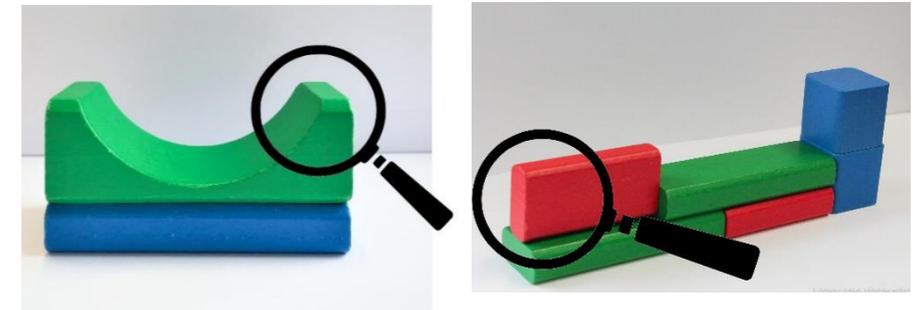


Bewertung Schreibprodukte?

Ver_rückte Gebäude – Subskala adäquate Informationsdichte

Konstruktionsprinzipien (Hennes et al., 2018):

- Jedes Gebäude enthält eine Lupenstelle (kritisches Element, das explizite Informationen erfordert)
- Bausteine variieren hinsichtlich Farbe & Form nur gemeinsam
- Steigende Komplexität der Bausteingebäude
- sachliches Vorwissen a priori im Test vorgegeben
- Adressatenkreis eindeutig benannt und Wissen explizit beschrieben



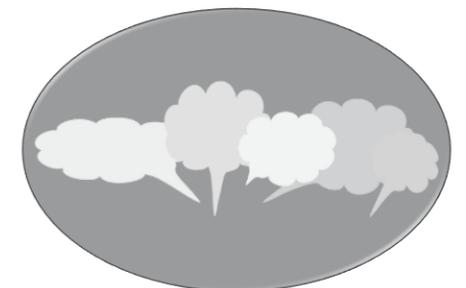
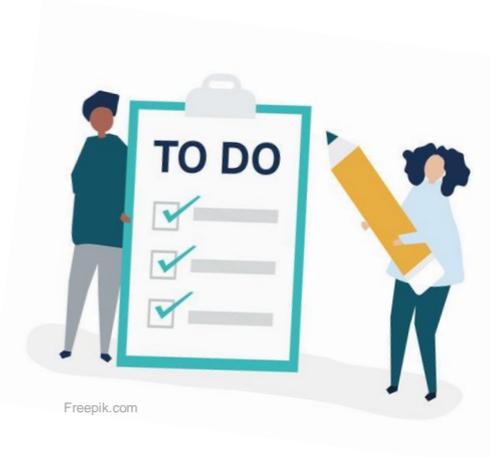
- Schau dir das Bild genau an. Baue das Gebäude nach.
- Überprüfe, ob du alles richtig gebaut hast. Achte besonders auf diese Stelle [Lupe].
- Ein Freund soll das Gebäude nachbauen. Er hat das Bild nicht. Schreibe ihm eine genaue Anleitung dazu, wie er das Gebäude nachbauen kann.

Anforderung: Minimum an notwendigen Informationen geben

Bewertung Schreibprodukte?

Sie sind wieder dran 😊: Breakout Session II

- Wie würden Sie die Schreibprodukte zu den beiden „ver_rückten Gebäuden“ bewerten? Diskutieren Sie!
- Entwickeln Sie ein Bewertungsraster, welches die Qualität des Schreibproduktes quantifiziert (z.B. in Form von Punkten)!
- Wenden Sie das erarbeitete Raster auf die Beispieltex-te an!
- Dauer der Breakout Session: 30 Minuten
- Anschließend Vorstellung Bewertungsraster und Vorgehen bei Erstellung pro Gruppe



Auswertung ver_rückte Gebäude

Entwicklung des Auswertungsrasters (Hennes, 2020)

- Versuchspersonen (N= 193 Studierende der Universität zu Köln, durchschnittliches Alter 22 Jahre, 85,5% weiblich) bekamen verschiedene Varianten von Bauanleitungen vorgegeben
- Informationsdichte in Bauanleitungen wurde durch Weglassen von Informationen variiert
- Überprüfung, ob der Nachbau trotz geringerer Informationsdichte noch möglich war
- Kriterium: 80 Prozent der VP konnten Gebäude anhand der gegebenen Informationen nachbauen

Auswertung ver_rückte Gebäude

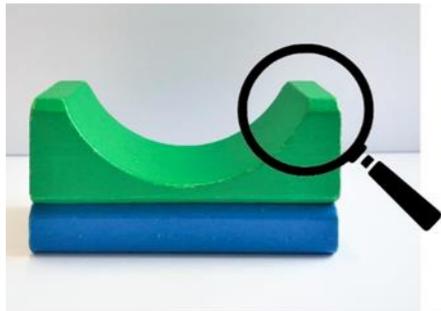
Score zur Informationsdichte

- Punktevergabe entlang Kriterienkatalog, der für jedes Gebäude/jede Schwierigkeitsstufe unbedingt notwendige Informationen vorgibt
- Für jede notwendige Information, die im Text gegeben wurde, wird ein Punkt vergeben → Quotient aus den gegebenen und allen notwendigen Informationen als Maß für die Informationsdichte (pro Schwierigkeitsstufe)

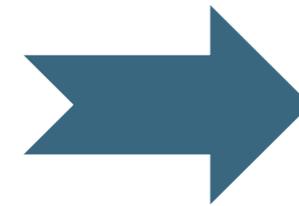
Auswertung ver_rückte Gebäude

Herausforderungen (Hennes, 2020)

- Was sind „Defaults“?
- Nicht alle Informationen müssen explizit gegeben werden



	gegeben	vs.	nicht gegeben
1) Blau liegt	<input type="checkbox"/>		<input checked="" type="checkbox"/>
2) Grün liegt über blau	<input checked="" type="checkbox"/>		<input type="checkbox"/>
3) Grün mit den Füßen nach oben	<input checked="" type="checkbox"/>		<input type="checkbox"/>



Unmögliches Antwortmuster, denn wenn grün über blau liegt, dann muss blau ebenfalls liegen -> implizite Information

Auswertung ver_rückte Gebäude

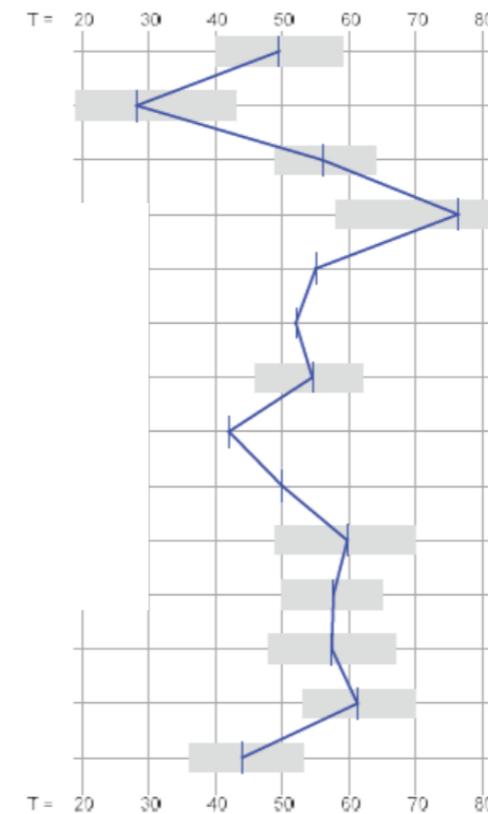
Herausforderungen (Hennes, 2020)

- Zulässige Formulierungsalternativen?
- 98 VP bewerten Adäquatheit von Formulierungen (N=106)
- Zusätzlich sahen die VP ein Foto des betreffenden Bausteins in seiner Zielposition sowie das Kriterium des Auswertungsrasters (z.B. „grün mit den Füßen nach oben“).
- Bestätigten mehr als 75% der VP die Funktionstüchtigkeit
→passende Formulierung

geht	geht nicht
Den grünen mit der Öffnung nach oben	Den grünen mit der Kuhle nach oben
Den grünen, wie eine Skaterbahn	Den grünen, wie eine Wippe
Die grüne Brücke falsch herum	Den grünen, wie eine Rampe
Den grünen mit der Einbuchtung nach oben	Den grünen auf die gerade Seite lege

Fazit

- Textbewertung nicht nur auf hierarchieniedriger Ebene!
- Textbewertung hierarchiehöherer Kompetenzen schwierig
- aber lohnenswert, da konkrete Ansätze möglich für Förderung relevanter und distinkter Fertigkeiten/Kompetenzen



Beispielhaftes
Kompetenzprofil

Literatur

- Bach, R., Schmidt, B.M., Schabmann, A. & van Koll, S. (2016). Braucht mein Friseur wirklich Zirkel und Lineal? – Schulisches Basiswissen im Kontext der Ausbildungsreife. *Heilpädagogische Forschung*, 42, 61-72.
- Baird, J. A., Greatorex, J. & Bell, J. F. (2004). What makes marking reliable? Experiments with UK examinations. *Assessment in education: Principles, policy & practice*, 11(3), 331-348.
- Becker-Mrotzek, M. (2014). Schreibkompetenz. In: J. Grabowski (Hrsg.), *Sinn und Unsinn von Kompetenzen: Fähigkeitskonzepte im Bereich von Sprache, Medien und Kultur*, 51-72. Opladen: Verlag Barbara Budrich.
- Beck, S. W., Llosa, L., Black, K., & Anderson, A. T. (2018). From assessing to teaching writing: What teachers prioritize. *Assessing Writing*, 37, 68-77. <https://doi.org/10.1016/j.asw.2018.03.003>

Literatur

- Birkel, P. & Birkel, C. (2002). Wie einig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht*, 49, 219-224.
- Brown, G. T., Glasswell, K. & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing writing*, 9(2), 105-121.
- Cooksey, R. W., Freebody, P. & Wyatt-Smith, C. (2007). Assessment as judgment-in-context: Analysing how teachers evaluate students' writing. *Educational Research and Evaluation*, 13(5), 401-434.
- Crossley, S. A. & McNamara, D. S. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, 7, 351-370.

Literatur

- Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25(2), 155-185.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445-476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Grice, H. P. (1975). Logic and conversation. In: P. Cole & J. Morgan (Hrsg.), *Syntax and Semantics: Speech Acts*, 3, 41-58. New York: University Press.
- Hamill, D.D. & Larsen, S.C. (2009). *Test of Written Language - TOWL (Fourth Edition)*. Austin, TX: Pro-ed.
- Hattie, J. A. C., & Learning, V. (2009). *A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Literatur

- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, 29, 369-388.
- Hennes, A. K. (2020). Schreibproduktebewerten – Die Rolle der Expertise bei der Bewertung der Textproduktionskompetenz. Doktorarbeit, Universität zu Köln. <https://kups.ub.uni-koeln.de/11414/>
- Hennes, A-K., Schmidt, B. M., Zepnik, S., Linnemann, M., Jost, J., Becker-Mrotzek, M., Rietz, C. & Schabmann, A. (2018). Schreibkompetenz diagnostizieren: Ein standardisiertes Testverfahren für die Klassenstufen 4-9 in der Entwicklung. *Empirische Sonderpädagogik*, 3, 294-310.
- Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik (Bd. 6)*. Weinheim: Beltz Verlag.
- Jølle, L. (2014). Pair assessment of pupil writing: A dialogic approach for studying the development of rater competence. *Assessing writing*, 20, 37-52. <https://doi.org/10.1016/j.asw.2014.01.002>

Literatur

- Kaufman, A.S. & Kaufman, N.L. (2005). Kaufman Test of Educational Achievement (Second Edition). London, UK: Pearson.
- Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16(2), 81-96.
- Langer, J. A. (2017). Introduction: Arthur N. Applebee: A Scholar's Life in Retrospect. In Durst, R. K., Newell, G. E., & Marshall, J. D. (Eds.). *English Language Arts Research and Teaching: Revisiting and Extending Arthur Applebee's Contributions*, 1-24. New York, London: Routledge.
<https://doi.org/10.4324/9781315465616-1>

Literatur

- Lehmann, R. H., Peek, R., & Poerschke, J. (2006). HAMLET 3-4: Hamburger Lesetest für 3. und 4. Klassen. Göttingen: Hogrefe.
- Lenhard, W., & Schneider, W. (2006). Ein Leseverständnistest für Erst- und Sechstklässler: ELFE 1-6. Göttingen: Hogrefe.
- Lim, G. S. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28, 543–560.
<https://doi.org/10.1177/0265532211406422>
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, 19(3), 246-276.
- MacArthur, C. A., Graham, S. & Fitzgerald, J. (2016). *Handbook of Writing Research*. New York: Guilford Press.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (2012). *Richtlinien und Lehrpläne für die Grundschule in Nordrhein-Westfalen*. Frechen: Ritterbach Verlag GmbH.

Literatur

- Rezaei, A. R. & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), 18-39.
- Schneider, W., Schlagmüller, M. & Ennemoser, M. (in press). Lesegeschwindigkeits- und -verständnistest für die Klassenstufen 5-13 (LGVT 5-12+). Göttingen: Hogrefe.
- Skar, G. B., & Jølle, L. J. (2017). Teachers as raters: Investigation of a long-term writing assessment program. *L1 Educational Studies in Language and Literature*, 17, 1-30. <https://doi.org/10.17239/L1ESLL-2017.17.01.06>
- Todd, R. W., Thienpermpool, P. & Keyuravong, S. (2004). Measuring the coherence of writing using topic-based analysis. *Assessing writing*, 9(2), 85-104.

Literatur

- Van Gasse, R., Lesterhuis, M., Verhavert, S., Bouwer, R., Vanhoof, J., Van Petegem, P., & De Maeyer, S. (2019). Encouraging professional learning communities to increase the shared consensus in writing assessments. *Journal of Professional Capital and Community*, 4(4), 269-285.
<https://doi.org/10.1108/JPC-08-2018-0021>
- Vögelin, C., Jansen, T., Keller, S. D. & Möller, J. (2018). The impact of vocabulary and spelling on judgments of ESL essays: An analysis of teacher comments. *The Language Learning Journal*, 1-17.
- Vögelin, C., Jansen, T., Keller, S. D., Machts, N. & Möller, J. (2019). The influence of lexical features on teacher judgements of ESL argumentative essays. *Assessing Writing*, 39, 50-63.
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: University Press.