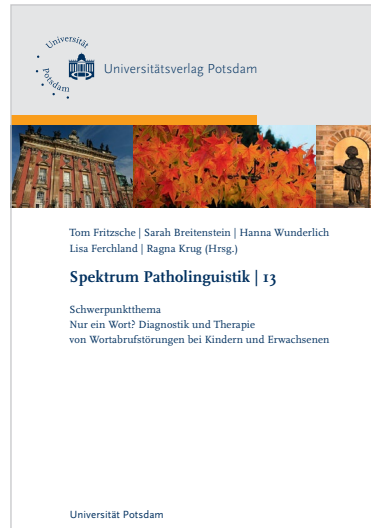


Artikel erschienen in:

*Tom Fritzsche, Sarah Breitenstein,
Hanna Wunderlich, Lisa Ferchland, Ragna Krug
(Hrsg.)*

Spektrum Patholinguistik Band 13. Schwerpunktthema: Nur ein Wort? Diagnostik und Therapie von Wortabruf- störungen bei Kindern und Erwachsenen

2020 – viii, 209 S.
ISBN 978-3-86956-488-3
DOI <https://doi.org/10.25932/publishup-46077>



Empfohlene Zitation:

Sophie Gruhn; Eliane Segers; Jos Keuning; Ludo Verhoeven: Dynamischer Leseverständnistest zur Differenzierung der Lernbedürfnisse von Grundschulkindern, In: Tom Fritzsche, Sarah Breitenstein, Hanna Wunderlich, Lisa Ferchland, Ragna Krug (Hrsg.): Spektrum Patholinguistik 13, Potsdam, Universitätsverlag Potsdam, 2020, S. 153–168.
DOI <https://doi.org/10.25932/publishup-47537>

Soweit nicht anders gekennzeichnet ist dieses Werk unter einem Creative Commons Lizenzvertrag lizenziert: Namensnennung 4.0. Dies gilt nicht für zitierte Inhalte anderer Autoren:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Dynamischer Leseverständnistest zur Differenzierung der Lernbedürfnisse von Grundschulkindern

Sophie Gruhn¹, Eliane Segers¹, Jos Keuning² & Ludo Verhoeven¹

¹ Behavioural Science Institute, Radboud University, Nijmegen

² Cito, Institute for Educational Measurement, Arnhem

1 Einleitung

Gewöhnliche Leseverständnistests geben keine Informationen über die dem Leseverständnis zugrundeliegenden Fähigkeiten (z. B. Dekodieren, Wortschatz, Satzintegrationen), welche zur Differenzierung der Lernbedürfnisse im Leseverständnis von Kindern in der Grundschule notwendig sind (Cain & Oakhill, 2012; Van den Broek, 2012). Diese Teilkomponenten werden häufig anhand einzelner Diagnostikverfahren erfasst und miteinander in Beziehung gesetzt, um die Lernbedürfnisse zu entschlüsseln. Diese isolierten Testverfahren berücksichtigen jedoch nicht, dass die Teilkomponenten im Leseverständnisprozess miteinander interagieren und geben daher nur ein eingeschränktes Bild auf die dem Leseverständnisprozess zugrundeliegenden Fähigkeiten (Perfetti & Adlof, 2012; Sabatini et al., 2016). Weiterhin geben herkömmliche Tests keine Auskunft über die Förderintensität, die zum aktuellen Zeitpunkt nötig ist, um die Leseentwicklung voranzutreiben. Der Unterricht wird oftmals erst dann angepasst oder intensiviert, wenn bereits ein Entwicklungsrückstand oder -stillstand festgestellt wurde (Compton et al., 2012; O'Reilly et al., 2012). Daher ist es wünschenswert, dass beim Erfassen der Lernbedürfnisse im Leseverständnis die Teilkomponenten in Interaktion miteinander gemessen werden und die nötige Intensität der Förderung im Unterricht im Voraus bestimmt werden kann.

Die genannten Ziele können durch eine dynamische Herangehensweise umgesetzt werden, bei der die Teilkomponenten im Leseverständnis auf Wort-, Satz- und Textniveau innerhalb desselben Textes erfasst werden sowie Feedback nach Fehlern gegeben wird

(Den Ouden et al., 2019). Diese zwei Aspekte wurden in vorherigen Testverfahren bisher jedoch noch nicht im vollen Ausmaß kombiniert (Den Ouden et al., 2019; Dörfler et al., 2017; Elleman et al., 2011; Navarro & Mora, 2011; Sabatini et al., 2015; 2016; 2020). Tests, die Instruktionen enthalten, sind dynamisch und können neben dem jetzigen Kenntnisstand auch das Lernpotential ermitteln (Sternberg & Grigorenko, 2002). Die Instruktion kann in Form eines Trainings zwischen einem Pretest und Posttest gegeben werden oder in Form von Feedback nach Fehlern direkt im Test (Dörfler et al., 2017). Das Lernpotential kann durch die Verbesserung vom Pretest zum Posttest ausgedrückt werden bzw. durch die Fähigkeit, nach dem Feedback eine richtige Antwort zu geben. Dies kann als Indikator für die nötige Förderintensität im Unterricht genutzt werden (Gustafson et al., 2014).

In dieser Studie wird davon ausgegangen, dass sich durch das Geben von Feedback zu den Teilkomponenten innerhalb eines Textes (z.B. Bedeutung von Wörtern und Integration von Sätzen im Text) das globale Verständnis des Lesers für diesen Text verbessert. Diese Annahme basiert auf den zahlreichen Studienergebnissen, die auf eine Verbesserung im Leseverständnis durch Wortschatz- und Satzintegrationsübungen hinweisen (z.B. Elleman, 2017; Elleman et al., 2009). Es ist jedoch unklar, inwiefern die Verbesserung im globalen Textverständnis sowie die Reaktion auf das Feedback zu den Teilkomponenten Hinweise auf das Lernpotential der Kinder im Leseverständnis liefern und somit zur Differenzierung der Lernbedürfnisse beitragen können. Bisher wurde noch nicht systematisch geprüft, ob sich das globale Verständnis eines Textes durch das Geben von Feedback auf Wortbedeutungen und Satzintegrationen innerhalb desselben Textes verbessert und was die isolierten und gemeinsamen Feedbackeffekte sind. In einer experimentellen Studie mit einem randomisiertem Pretest-Posttest Kontrollgruppendesign wurde daher untersucht, wie sich das globale Textverständnis von Dritt- bis Fünftklässler*innen durch das Geben von Feedback zu Wortbedeutungs- und Satzintegrationsfragen oder nur zu Wortbedeutungs-

fragen innerhalb eines Textes vom Pretest zum Posttest verbessert. Weiterhin wurde überprüft, ob die Wortbedeutungskennnisse und Satzintegrationsfähigkeiten ohne Feedback und die Reaktion auf das Feedback zur Differenzierung von unterschiedlichen Leseverständnissfähigkeiten beitragen können. Dazu wurde ermittelt, ob die Verbesserung im globalen Verständnis eines Textes von den Antworten zu Wortbedeutungs- und Satzintegrationsfragen innerhalb desselben Textes beim ersten Versuch und nach Feedback abhängt.

2 Methode

2.1 Teilnehmer*innen

An dieser Studie haben 293 Dritt- bis Fünftklässler*innen einer vorstädtischen Gemeinde im Osten der Niederlande im gleichen Zeitraum teilgenommen. Die Teilnehmer*innen sprachen Niederländisch als Erstsprache oder als Zweitsprache und wurden innerhalb der Klassen zufällig der Wortbedeutungsgruppe ($n = 97$), Wortbedeutungs- und Satzintegrationsgruppe ($n = 97$) oder Kontrollgruppe zugeordnet ($n = 98$).

2.2 Material

Es wurde der dynamische computerbasierte Leseverständnistest aus der Studie von Den Ouden und Kollegen (2019) verwendet. Der Test bestand aus 30 kurzen niederländischen Texten mit jeweils vier Fragen: eine globale Textverständnisfrage, zwei Wortbedeutungsfragen und eine Satzintegrationsfrage. Die Texte waren entweder narrativ oder informativ. Die Themen gehörten nicht zum Allgemeinwissen von Kindern und waren altersentsprechend formuliert. Die Textlänge war durchschnittlich 164,7 Wörter ($SD = 44,4$) für Drittklässler*innen,

182,6 Wörter ($SD = 78,0$) für Viertklässler*innen und 205,2 Wörter ($SD = 44,6$) für Fünftklässler*innen. Alle Fragen wurden im Multiple-Choice-Format mit drei oder vier Antwortmöglichkeiten präsentiert. Pro Klassenstufe wurde eine andere Testversion vorgelegt, wobei der Schwierigkeitsgrad der Texte und der Inhalt an die Klassenstufe angepasst wurde.

2.3 Durchführung

Der dynamische Leseverständnistest wurde in sechs Teiltests von jeweils fünf Texten innerhalb von sechs Wochen klassenweise abgenommen. Jeder Teiltest bestand aus einem Pretest und einem Posttest. Pro Teiltest wurden erst der Pretest und der Posttest abgeschlossen, bevor zum nächsten Teiltest übergegangen wurde (d.h. Pretest Teiltest 1, Posttest Teiltest 1, Pretest Teiltest 2, Posttest Teiltest 2, ..., Pretest Teiltest 6, Posttest Teiltest 6). Zwischen jedem Pretest und Posttest eines Teiltests lagen drei Tage. An einem Tag konnte nur ein Pretest oder ein Posttest durchgeführt werden.

Pretest. Die Teilnehmer*innen lasen pro Teiltest fünf Texte. Nach jedem Text beantworteten sie eine globale Textverständnisfrage. Es wurde kein Feedback zur Korrektheit der Antworten gegeben. Der Pretest unterschied sich nicht zwischen der Wortbedeutungs- und Satzintegrationsgruppe, Wortbedeutungsgruppe und Kontrollgruppe.

Posttest. Die Teilnehmer*innen lasen dieselben Texte und beantworteten dieselbe globale Textverständnisfrage wie beim Pretest. Es wurde Feedback zur Korrektheit der Antworten gegeben. Abhängig von der Gruppenzugehörigkeit (Wortbedeutungs- und Satzintegrationsgruppe, Wortbedeutungsgruppe, Kontrollgruppe) wurden beim Posttest pro Text zusätzliche Fragen vor der globalen Textverständnisfrage gestellt.

- *Wortbedeutungs- und Satzintegrationsgruppe:* Die Wortbedeutungs- und Satzintegrationsgruppe beantwortete pro Text zusätzlich zwei Wortbedeutungsfragen mit Feedback und eine Satzintegrationsfrage mit Feedback.
- *Wortbedeutungsgruppe:* Die Wortbedeutungsgruppe erhielt pro Text nur die zwei Wortbedeutungsfragen mit Feedback.
- *Kontrollgruppe:* Die Kontrollgruppe beantwortete nur die globalen Textverständnisfragen und keine zusätzlichen Fragen.

2.4 Feedback zu Wortbedeutungs- und Satzintegrationsfragen

Im Posttest wurde in den experimentellen Gruppen (Wortbedeutungs- und Satzintegrationsgruppe, Wortbedeutungsgruppe) Feedback zur Korrektheit der Antworten auf die Wortbedeutungs- und Satzintegrationsfragen gegeben. Nach falschen Antworten konnte erneut eine Antwort gegeben werden. Hierbei wurde eine inhaltliche Hilfe präsentiert. Wenn die Antwort beim zweiten Versuch erneut falsch war, wurde die richtige Antwort vorgegeben.

3 Hypothesen und Analyse

Aufgrund des binären Datensatzes (korrekt vs. falsch) wurde mit generalisierten gemischten linearen Modellen untersucht, wie sich die Wahrscheinlichkeit, die globale Textverständnisfrage eines Textes korrekt zu beantworten, vom Pretest zum Posttest verändert (Bates et al., 2015). Dazu wurde pro Forschungsfrage die Korrektheit der globalen Textverständnisfrage eines Textes (korrekt vs. falsch) als abhängige Variable und der Faktor Zeit (Levels: Pretest, Posttest) als fester Haupteffekt spezifiziert. Dieses „Basismodell“ wurde pro Forschungsfrage mit weiteren festen Effekten ergänzt, um zu untersuchen, wie die jeweiligen experimentellen Faktoren die Korrektheit

der Antworten auf die globale Textverständnisfrage beeinflussen. Diese weiteren festen Effekte werden im Folgenden pro Forschungsfrage spezifiziert. Pro Modell wurden zufällige Abschnitte für Teilnehmer*innen und Texte genutzt sowie zufällige Anstiege für die experimentellen Prädiktoren, sofern diese den Modellfit signifikant verbesserten (Barr et al., 2013). Die Modelle wurden jeweils mit sum-to-zero Kontrasten geschätzt. Mit Post-hoc-Tests wurde die Signifikanz zwischen den unterschiedlichen Faktorlevels untersucht.

3.1 Unterschiede zwischen den Untersuchungsbedingungen

3.1.1 *Hypothese*

Es wurde davon ausgegangen, dass das Feedback zu einer Verbesserung im globalen Textverständnis vom Pretest zum Posttest führt (Elleman, 2017; Elleman et al., 2009). Daher wurde in den experimentellen Gruppen ein signifikant stärkerer Anstieg vom Pretest zum Posttest in der Wahrscheinlichkeit, die globale Textverständnisfrage eines Textes korrekt zu beantworten (d.h. eine Verbesserung im globalen Textverständnis), erwartet als in der Kontrollgruppe. Weiterhin wurde angenommen, dass sich die Wortbedeutungs- und Satzintegrationsgruppe stärker im globalen Textverständnis verbessert als die Wortbedeutungsgruppe, da Feedback auf mehrere Teilkomponenten vermutlich einen stärkeren Effekt hat als nur auf eine Teilkomponente. In der Kontrollgruppe wurde keine Veränderung im globalen Textverständnis erwartet.

3.1.2 *Analyse*

Neben dem Haupteffekt Zeit (Kontraste pro Faktorlevel: Pretest = -1, Posttest = 1) wurde der Faktor Gruppe (Kontraste pro Faktorlevel: Wortbedeutungs- und Satzintegrationsgruppe = 1, 0; Wortbedeutungsgruppe = 0, 1; Kontrollgruppe = -1, -1) als Haupteffekt sowie deren Zweiweginteraktion zum Modell hinzugefügt. Mit Post-hoc-Tests wurde überprüft, inwiefern sich die Gruppen signifikant voneinander unterschieden bzw. sich signifikant vom Pretest zum Posttest verbesserten.

3.2 Moderation durch Feedback

3.2.1 *Hypothese*

Eine Verbesserung vom Pretest zum Posttest im globalen Textverständnis wurde nur erwartet, wenn Feedback für die Wortbedeutungs- und Satzintegrationsfragen benötigt wurde (d. h. die schwächere Gruppe), aber nicht, wenn kein Feedback benötigt wurde (d. h. die stärkere Gruppe). Weiterhin wurde eine stärkere Verbesserung im globalen Textverständnis angenommen, wenn nach dem Feedback eine richtige Antwort auf die Wortbedeutungs- und Satzintegrationsfragen gegeben wurde (d. h. die mittelstarke Gruppe), als wenn eine falsche Antwort gegeben wurde (d. h. die schwächste Gruppe). Dies wurde für die Wortbedeutungs- und Satzintegrationsgruppe sowie Wortbedeutungsgruppe gleichermaßen angenommen.

3.2.2 *Analyse*

Die Analyse wurde separat für die Wortbedeutungs- und Satzintegrationsfragen sowie separat für die jeweilige Korrektheit der Antworten

auf die Wortbedeutungs- und Satzintegrationsfragen beim ersten oder zweiten Versuch ausgeführt. Dies ergab vier Modelle. Neben dem Haupteffekt Zeit (Kontraste pro Faktorlevel: Pretest = -1 , Posttest = 1) wurde als Haupteffekt der Faktor Korrektheit der Antworten beim ersten Versuch (Kontraste pro Faktorlevel: korrekt = 1 , falsch = -1) bzw. der Faktor Korrektheit der Antworten beim zweiten Versuch (Kontraste pro Faktorlevel: korrekt = 1 , falsch = -1) hinzugefügt sowie deren jeweilige Zweiweginteraktion mit Zeit. Mit Post-hoc-Tests wurde untersucht, inwiefern die Veränderung im globalen Textverständnis vom Pretest zum Posttest pro Kontrastlevel (korrekt, falsch) signifikant war. Bei der Analyse zu den Wortbedeutungsfragen, wurde die Gruppenzugehörigkeit als Kontrollvariable hinzugefügt (Kontraste pro Faktorlevel: Wortbedeutungs- und Satzintegrationsgruppe = 1 , Wortbedeutungsgruppe = -1). Bei der Analyse zu den Satzintegrationsfragen wurde für die Antworten auf die Wortbedeutungsfragen kontrolliert. Dazu wurde pro Wortbedeutungsfrage die Korrektheit der Antwort kontinuierlich bewertet (2 = korrekt ohne Feedback, 1 = korrekt nach Feedback, 0 = falsch nach Feedback) und pro Text für die zwei Items summiert.

4 Ergebnisse

Die Ergebnisse wurden in diesem Beitrag nur deskriptiv dargestellt. Eine ausführliche Beschreibung wird in einer wissenschaftlichen Fachzeitschrift veröffentlicht.

4.1 Unterschiede zwischen den Untersuchungsbedingungen

Da die Interaktion zwischen Zeit und Gruppe signifikant war, unterschieden sich die Gruppen in der Veränderung des globalen Textverständnisses vom Pretest zum Posttest. Entsprechend der Erwartung

gab es in der Kontrollgruppe keine signifikante Veränderung im globalen Textverständnis. Jedoch verschlechterten sich im Gegensatz zu unseren Erwartungen beide experimentellen Gruppen vom Pretest zum Posttest im globalen Textverständnis. Diese Verschlechterung war gegenüber der Kontrollgruppe nur für die Wortbedeutungs- und Satzintegrationsgruppe signifikant. Die experimentellen Gruppen unterschieden sich nicht voneinander.

4.2 Explorative Untersuchung: Moderation durch Feedback

Obwohl die Ergebnisse der zuvor beschriebenen Analyse unseren Erwartungen widersprachen, wurde die Analyse zur zweiten Forschungsfrage wie geplant ausgeführt. Dadurch konnte exploriert werden, inwiefern das Antwortverhalten der Teilnehmer*innen Hinweise auf die Ursachen in der Verschlechterung im globalen Textverständnis liefern kann. Die Analyse zeigte, dass die oben beschriebene Verschlechterung im globalen Textverständnis von der Korrektheit der Antworten auf die Wortbedeutungs- und Satzintegrationsfragen beim ersten sowie beim zweiten Versuch abhängig war. Die Zweiweginteraktion zwischen Zeit und der Korrektheit der Antworten beim ersten Versuch bzw. beim zweiten Versuch waren pro Fragentyp signifikant. Die Post-hoc-Tests zeigten, dass das globale Textverständnis sich vom Pretest zum Posttest nicht veränderte, wenn die Wortbedeutungsfrage oder Satzintegrationsfrage zu einem Text beim ersten Versuch korrekt beantwortet wurde. Eine Verschlechterung im globalen Textverständnis trat nur auf, wenn Feedback benötigt wurde. Wiederum war die Verschlechterung stärker, wenn nach dem Feedback erneut eine falsche Antwort gegeben wurde, als wenn die Antwort beim zweiten Versuch korrekt war.

5 Schlussfolgerungen

5.1 Unterschiede zwischen den Untersuchungsbedingungen

Die experimentellen Gruppen, die Fragen und Feedback zu den Teilkomponenten innerhalb eines Textes empfangen, verschlechterten sich im globalen Verständnis der Texte, während sich die Kontrollgruppe ohne diese Fragen und Feedback nicht im globalen Textverständnis veränderte. Dies war unerwartet, da zahlreiche Studien positive Effekte von Wortschatz- und Satzintegrationsinterventionen auf das Leseverständnis fanden (Elleman, 2017; Elleman et al., 2009). Da die Wortbedeutungs- und Satzintegrationsgruppe sich stärker verschlechterte als die Wortbedeutungsgruppe, schien die Verschlechterung im globalen Textverständnis mit der Anzahl der zu beantwortenden Fragen beim Posttest zusammenzuhängen und war daher vermutlich auf eine Erschöpfung und kognitive Überlastung zurückzuführen. Weiterhin schienen die zusätzlichen Fragen und das Feedback ungenügend in den Leseprozess integriert zu sein und könnten somit vom globalen Verständnis der Texte abgelenkt haben (Al-Shehri & Gitsaki, 2010; Marefat et al., 2016). Beispielsweise wurde als Feedback zu falsch beantworteten Wortbedeutungsfragen, Bilder präsentiert, welche die Wortbedeutung darstellten, jedoch nicht auf den Textinhalt abgestimmt waren (Carney & Levin, 2002). Ein weiterer Aspekt kann auch die fehlende Vertrautheit mit dem Testformat in den experimentellen Gruppen bezüglich des Empfangens von Feedbacks sein (Sabatini et al., 2014), während die Kontrollgruppe gewöhnliche Tests ohne Feedback und zusätzliche Fragen erhielt.

5.2 Moderation durch Feedback

Die Ergebnisse zur Verschlechterung im globalen Textverständnis konnten weiter spezifiziert werden, da dies mit der Korrektheit der Antworten auf die Wortbedeutungs- und Satzintegrationsfragen beim ersten und zweiten Versuch zusammenhing. Für die stärkeren Schüler*innen, die kein Feedback benötigten, blieb das globale Textverständnis wie in der Kontrollgruppe konstant. Nur die schwächeren Schüler*innen, welche Feedback benötigten, verschlechterten sich im globalen Textverständnis. Diese Verschlechterung war stärker, wenn nach dem Feedback erneut eine falsche Antwort gegeben wurde, d. h. für die schwächsten Schüler*innen. Dies könnte unter anderem damit erklärt werden, dass vor allem Schüler*innen mit geringeren Lesefähigkeiten wahrscheinlich häufiger Feedback benötigten. Durch das erneute Beantworten der Fragen beim zweiten Versuch sowie das Verarbeiten des Feedbacks waren sie womöglich einer stärkeren kognitiven Belastung ausgesetzt, zusätzlich zu den sowieso schon geringeren Leseverständnisfähigkeiten. Schüler*innen mit geringeren Fähigkeiten in den Teilkomponenten (Wortschatz und Satzintegration) haben vermutlich ein weniger stabiles globales Textverständnis beim Pretest und sind daher anfälliger für die zu der Verschlechterung führenden Effekte (Perfetti & Adolf, 2012). Somit könnte die Verschlechterung im globalen Textverständnis mit unterschiedlichen Niveaus im Leseverständnis im Zusammenhang stehen, die durch das Antwortverhalten bei den Wortbedeutungs- und Satzintegrationsfragen unterschieden werden können.

5.3 Limitationen und zukünftige Forschungsfragen

Eine Limitation der Studie ist die fehlende Vergleichbarkeit der kognitiven Belastung sowie Beschäftigungsdauer mit der Aufgabe beim Posttest zwischen den Untersuchungsbedingungen. Eine weitere Limitation ist, dass keine Aussagen zur Effektivität des Feedbacks auf

Wortbedeutungs- und Satzintegrationsfragen gemacht werden können. Zukünftige Studien könnten eine oder mehrere Kontrollgruppen einsetzen, die die gleichen Fragen wie die experimentelle Gruppe aber ohne Feedback bzw. andere nicht auf den textbezogene Fragen beantworten. Außerdem sollte Feedback verwendet werden, das mehr in den Leseprozess integriert ist.

5.4 Fazit

Entgegen den Erwartungen führte Feedback beim Lesen eines Textes zu einer Verschlechterung im globalen Textverständnis, vermutlich aufgrund einer kognitiven Überlastung oder Ablenkung vom Textinhalt. Das Design des Tests eignete sich nicht, um das Lernpotential im globalen Textverständnis zu messen. Allerdings schien die Fähigkeit, die Wortbedeutungs- und Satzintegrationsfragen innerhalb eines Textes beim ersten Versuch bzw. mithilfe von Feedback korrekt zu beantworten, nicht nur individuelle Unterschiede in den Teilkomponenten aufzeigen zu können, sondern auch mit der Stabilität im globalen Textverständnis zusammen zu hängen. Die Notwendigkeit Feedback zu den Teilkomponenten innerhalb eines Textes zu erhalten sowie die Reaktion auf Feedback könnten demnach Hinweise auf die individuellen Unterschiede in den Leseverständnissfähigkeiten geben und somit die Lernbedürfnissen aufzeigen. Anhand der Häufigkeit des gegebenen Feedbacks zu Wortbedeutungen und Satzintegrationen, können Schüler*innen identifiziert werden, die vor allem in einer der beiden Teilkomponenten oder gleichermaßen in beiden Teilkomponenten Probleme haben. Außerdem kann die Fähigkeit nach dem Feedback korrekt zu antworten, zur Unterscheidung von Schüler*innen innerhalb des gleichen Lernschwerpunkts bezüglich der Förderungsintensität genutzt werden (z.B. Gustafson et al., 2014). Ein Test bei dem mehrere Teilkomponenten im Leseverständnis innerhalb eines Textes gemessen werden und Feedback nach Fehlern gegeben wird, könnte dazu beitragen, die bestehenden Probleme in

der Identifizierung von individuellen Lernbedürfnissen zu lösen und somit den Leseverständnisunterricht effizienter zu gestalten.

6 Literatur

- Al-Shehri, S. & Gitsaki, C. (2010). Online reading: A preliminary study of the impact of integrated and split-attention formats on L2 students cognitive load. *ReCALL*, 22 (3), 356–375. <https://doi.org/10.1017/S0958344010000212>
- Barr, D.J., Levy, R., & Scheepers, C. & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68 (3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B.M. & Walker, S.C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67 (1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Cain, K. & Oakhill, J. (2012). Reading comprehension development from seven to fourteen years: Implications for assessment. In J.P. Sabatini, E.R. Albro & T. O'Reilly (Hrsg.), *Measuring up: Advances in how to assess reading ability* (59–75). Rowman & Littlefield Education.
- Carney, R.N. & Levin, J.R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14 (1), 5–26. <https://doi.org/10.1023/A:1013176309260>
- Compton, D.L., Gilbert, J.K., Jenkins, J.R., Fuchs, D., Fuchs, L.S., Cho, E., Barquero, L.A., & Bouton, B. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities*, 45(3), 204–216. <https://doi.org/10.1177/0022219412442151>

- Den Ouden, M., Keuning, J. & Eggen, T. (2019). Fine-grained assessment of children's text comprehension skills. *Frontiers in Psychology, 10* (1313), 1–12, <https://doi.org/10.3389/fpsyg.2019.01313>
- Dörfler, T., Golke, S. & Artelt, C. (2017). Evaluating prerequisites for the development of a dynamic test of reading competence: Feedback effects on reading comprehension in children. In D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (Hrsg.), *Competence assessment in education: Research, models and instruments. Methodology of educational measurement and assessment* (487–503). Springer.
- Elleman, A.M. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review. *Journal of Educational Psychology, 109* (6), 761–781. <https://doi.org/10.1037/edu0000180>
- Elleman, A.M., Compton, D.L., Fuchs, D., Fuchs, L.S. & Bouton, B. (2011). Exploring dynamic assessment as a means of identifying children at risk of developing comprehension difficulties. *Journal of Learning Disabilities, 44* (4), 348–357. <https://doi.org/10.1177/0022219411407865>
- Elleman, A.M., Lindo, E.J., Morphy, P. & Compton, D.L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2* (1), 1–44. <https://doi.org/10.1080/19345740802539200>
- Gustafson, S., Svensson, I. & Fälth, L. (2014). Response to intervention and dynamic assessment: Implementing systematic, dynamic and individualised interventions in primary school. *International Journal of Disability, Development and Education, 61* (1), 27–43. <https://doi.org/10.1080/1034912X.2014.878538>

- Marefat, H., Rezaee, A.A. & Naserieh, F. (2016). Effect of computerized gloss presentation format on reading comprehension: A cognitive load perspective. *Journal of Information Technology Education: Research, 15*, 479–501. <https://doi.org/10.28945/3568>
- Navarro, J.-J. & Mora, J. (2011). Analysis of the implementation of a dynamic assessment device of processes involved in reading with learning-disabled children. *Learning and Individual Differences, 21* (2), 168–175. <https://doi.org/10.1016/j.lindif.2010.11.008>
- O'Reilly, T., Sabatini, J., Bruce, K., Pillarisetti, S. & McCormick, C. (2012). Middle school reading assessment: Measuring what matters under a RTI framework. *Reading Psychology, 33*(1-2), 162–189. <https://doi.org/10.1080/02702711.2012.631865>
- Perfetti, C. & Adlof, S.M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J.P. Sabatini, E.R. Albro & T. O'Reilly (Hrsg.), *Measuring up: Advances in how to assess reading ability* (3–20). Rowman & Littlefield Education.
- Sabatini, J., Bruce, K., Steinberg, J. & Weeks, J. (2015). SARA reading components tests, RISE forms: Technical adequacy and test design, 2nd edition. *ETS Research Report Series, 2*, 1–20. <https://doi.org/10.1002/ets2.12076>
- Sabatini, J.P., Halderman, L.K., O'Reilly, T. & Weeks, J.P. (2016). Assessing comprehension in kindergarten through third grade. *Topics in Language Disorders, 36* (4), 334–355. <https://doi.org/10.1097/TLD.0000000000000104>
- Sabatini, J.P., O'Reilly, T., Halderman, L. & Bruce, K. (2014). Broadening the scope of reading comprehension using scenario-based assessments: Preliminary findings and challenges. *L'Année psychologique, 114* (4), 693–723. <https://doi.org/10.4074/S0003503314004059>

- Sabatini, J., O'Reilly, T., Weeks, J. & Wang, Z. (2020). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*, 20(1) 1–23. <https://doi.org/10.1080/15305058.2018.1551224>
- Sternberg, R.J. & Grigorenko, E.L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge University Press.
- Van den Broek, P. (2012). Individual and developmental differences in reading comprehension: Assessing cognitive processes and outcomes. In J.P. Sabatini, E.R. Albro & T. O'Reilly (Hrsg.), *Measuring up: Advances in how to assess reading ability* (39–57). Rowman & Littlefield Education.

Danksagung

Diese Studie war Teil eines gemeinsamen Forschungsprojekts der Radboud University (Ludo Verhoeven, Eliane Segers, Sophie Gruhn), Cito und Twente University (Theo Eggen, Jos Keuning, Marije den Ouden), Expertisecentrum Nederlands (Nicole Heister-Swart) und Kennisinstituut voor Taalontwikkeling (ITTA, Femke Scheltinga). Wir bedanken uns bei The Netherlands Initiative for Education Research (National Regieorgaan Onderwijsonderzoek, NWO 405-15-548) für die Finanzierung des Forschungsprojektes.

Kontakt

Sophie Gruhn
c.gruhn@bsi.ru.nl