



Nikola Ebenbeck



Jana Jungjohann



Andreas Mübling



Markus Gebhardt

Die Bearbeitungsgeschwindigkeit von Kindern mit Lernschwierigkeiten als Grundlage für die Testentwicklung von Lernverlaufsdiagnostik

Lernverlaufsdiagnostik soll reliabel und möglichst kurz sein, um im Unterricht ökonomisch einsetzbar zu sein. Ziel eines kurzen Tests ist es, dass Schülerinnen und Schüler mit Lernschwierigkeiten im Testzeitraum zur Informationsgewinnung ausreichend Items bearbeiten. Die Bearbeitungsgeschwindigkeit ist hierfür ausschlaggebend. Der vorliegende Beitrag untersucht die Bearbeitungsgeschwindigkeit und -genauigkeit auf Item- und Testebene der – als Speedtests konstruierten – Lesetests der Onlineplattform Levumi.de von Schülerinnen und Schülern mit und ohne sonderpädagogischen Unterstützungsbedarf. Die Ergebnisse zeigen, dass Schülerinnen und Schüler mit Unterstützungsbedarf die Items im Schnitt langsamer und fehlerhafter bearbeiten als Schülerinnen und Schüler ohne Unterstützungsbedarf trotz eines höheren Durchschnittsalters zum Testzeitpunkt. Es wird diskutiert, welche Schlüsse auf Grundlage der Bearbeitungsgeschwindigkeit gezogen werden können und inwiefern sich „gute Testergebnisse“ als Richtwerte für Lehrkräfte ableiten lassen.

Zusammenfassung

Pädagogische Diagnostik gewinnt im inklusiven Unterricht zunehmend an Bedeutung. Mit der Feststellung des aktuellen Lernstands und einer Dokumentation des Lernerfolgs kann Chancengleichheit und -gerechtigkeit sichergestellt werden. So fordert Artikel 24e UN-BRK den Nachweis der Wirksamkeit von Unterstützungsmaßnahmen. Für die Sonderpädagogik bedeutet das, dass die Bewertung und Evaluation von Unterricht und Maßnahmen einen höheren Stellenwert als bislang einnehmen. Zur Bewertung des schulischen Erfolgs aller Schülerinnen und Schüler benötigt man sowohl die Beobachtung der Lernausgangslage als auch der Lernentwicklung. In beiden Fällen können mittels standardisierter Tests Ergebnisse und Informationen ökonomisch gesammelt und anschließend ausgewertet werden. Die Erhebung der Lernausgangslage wird genutzt, um passende didaktische Methoden auszuwählen und zu begründen (Gebhardt & Jungjohann, 2020). Hierfür wird auf Statusdiagnostik zurückgegriffen, um ein differenziertes Profil der Schülerin oder des Schülers zu bestimmen. Dieses Vorgehen ist in Deutschland fest im schulischen Alltag in Form von Lernstandserhebungen oder zur Feststellung des sonderpädagogischen Unterstützungsbedarfs verankert (Gebhardt, Jungjohann & Schurig, 2021; Hartke, Sikora & Wember, 2021). Zur Messung der Lernentwicklung werden mithilfe sogenannter Lernverlaufsdiagnostik individuelle unterrichtsimmanente Lernverläufe in einem längeren Zeitraum dargestellt (Klauer, 2011; Gebhardt, Diehl & Mübling, 2015). Dabei kann sowohl der Erwerb von Kompetenzen als auch Fertigkeiten verfolgt werden (Jungjohann, Diehl, Mübling & Gebhardt, 2018). Lernverlaufsdiagnostik hat ihre Wurzeln im amerikanischen sonderpädagogischen Ansatz des Curriculum-Based-Measurement (Deno, 1985) und wird erst seit der Jahrtausendwende in Deutschland erforscht sowie im Unterricht eingesetzt.

Lernverlaufdiagnostik

Pädagogische Diagnostik soll ökonomisch, also einfach und schnell, einsetzbar sein. Statustests umfassen meist viele Items zu mehreren Kompetenzbereichen, um möglichst umfangreiche und detaillierte Informationen zu erhalten. Dadurch ist die Testzeit für die Schülerinnen und Schüler vergleichsweise hoch und kann bis zu einer Schulstunde betragen. Die Ökonomie ist demzufolge ein Nebengütekriterium in der Statusdiagnostik. In der Lernverlaufdiagnostik ist die Ökonomie dagegen das zentrale Gütekriterium. Durchschnittlich zeigen zwischen 10 und 30 % der Schülerinnen und Schüler keinen oder nur einen geringen Lernfortschritt im Unterricht (Anderson, Jungjohann & Gebhardt, 2020; Fuchs, Fuchs & Compton, 2004).

Eine Herausforderung in der pädagogischen Diagnostik ist, dass Schülerinnen und Schüler mit Lernschwierigkeiten oder einem erhöhten Risiko meist eine kürzere Konzentrationsspanne sowie Defizite im Arbeitsgedächtnis aufweisen (Tourva, Spanoudis & Demetriou, 2016; Schuchardt, Gebhardt & Mähler, 2010). Betroffene Kinder und Jugendliche bearbeiten Tests in einem vergleichbar höheren Alter meist langsamer und fehlerhafter (Müller et al., 2017; Klicpera & Gasteiger-Klicpera, 1993). Aus diesem Grund ist Lernverlaufdiagnostik für solche Kinder gut geeignet, da mittels kurzer Tests häufiger gemessen wird. Im Gegenteil dazu werden bei der Statusdiagnostik viele Informationen innerhalb einer langen Testung gewonnen (Vaughn et al., 2003). Lernverlaufstests dauern nur wenige Minuten. Daher können Schülerinnen und Schülern mit schwacher Konzentration ihre Kompetenzen ohne Konzentrationsverlust in solchen Tests besser zeigen.

Fest in den Schulalltag integriert kann Lernverlaufdiagnostik neben der Messung der Lernentwicklung auch zur Begründung und Planung der unterrichtlichen Förderung genutzt werden (Jungjohann, 2022). Dafür interpretieren die Lehrpersonen die Messergebnisse der Lernverlaufstests, verknüpfen die unterrichtliche Förderung mit den Lernentwicklungen und ziehen Rückschlüsse für den weiteren Unterricht (Jungjohann et al., 2018). Lernverlaufdiagnostik ist eng mit dem amerikanischen Response-To-Intervention-Ansatz verknüpft. Dieser Ansatz ist innerhalb der Mehrebenenprävention gerahmt. Dieses Modell verbindet evidenzbasierte und datengeleitete Praxis, betrachtet dabei sonderpädagogischen Unterstützungsbedarf präventiv und integrativ, und wird meist als Schulmodell mit mehreren Ebenen umgesetzt (Gresham, Van Der Heyden & Witt, 2005). Lernverlaufdiagnostik spielt eine wesentliche Rolle in der Förderplanung im Rügener Inklusionsmodell, indem Schülerinnen und Schüler mit Schwierigkeiten, z.B. im Lesen, Rechnen und in der Konzentration, zusätzliche Differenzierung im Unterricht (Förderebene II) oder intensive sonderpädagogische Förderung (Förderebene III) erhalten (Voß & Hartke, 2014). Wenn die Lehrpersonen in den Lernverlaufdaten in einem angemessenen Übungszeitraum keine ausreichenden Lernfortschritte beobachten und weitere Gründe für das Ausbleiben des Lernerfolgs ausgeschlossen sind, kann daraufhin die Förderung angepasst werden. Lehrpersonen nutzen die diagnostischen Daten einerseits, um Adaptionen im Unterricht zu begründen und andererseits, um Lernziele festzulegen.

Der Ansatz der Lernverlaufdiagnostik hat Konsequenzen für die testtheoretische Konstruktion. Die einzelnen Items sind entweder nach dem Prinzip des Robusten Indikators oder des Curriculum Samplings konstruiert (Fuchs, 2004). Robuste Indikatoren stehen stellvertretend für eine allgemeine Kompetenz, die mit der geprüften Kompetenz hoch korreliert. Durch die Messung des Robusten Indikators wird daher die jeweilige verknüpfte Kompetenz geprüft (Hartke et al., 2021). Beim Ansatz des Curriculum Samplings werden dagegen für einen Test unterschiedliche Aufgabentypen entwickelt, um mehrere Teilfähigkeiten einer Kompetenz in einem Test abzubilden. Lernverlaufdiagnostik kann dabei entweder als Paper-Pencil-Test oder als Computertest realisiert werden (Fuchs, Fuchs, Hamlett & Stecker, 1990). In Deutschland ist dies beispielsweise durch die Verfahren von Lernlinien, Levumi oder quop möglich (siehe für einen Vergleich: Blumenthal, Gebhardt, Förster & Souvignier, 2022).

Unabhängig von der Aufgabenart sind die Lernverlaufstests häufig als Speedtests (van Breukelen, 2005) konstruiert. Die Schülerinnen und Schüler bearbeiten in einer festgelegten Zeitspanne so viele Items wie möglich. Die Items sind dabei alle vergleichbar im Schwierigkeitsgrad. Um den Erfolg im jeweiligen Test einschätzen zu können, ist die Anzahl an richtig gelösten Items inner-

halb der Zeitspanne relevant. Je nach Bearbeitungsgeschwindigkeit lösen die Schülerinnen und Schüler also eine unterschiedliche Anzahl an Items in demselben Test. Eine andere Möglichkeit ist die Konstruktion als Powertests. Hier wird nicht die Geschwindigkeit, sondern die Leistungshöhe gemessen. Die Items sind unterschiedlich schwer und werden nach Schwierigkeit geordnet. Mit dem festgelegten Zeitlimit kann die Lehrkraft ihren Unterricht oder die Freiarbeitszeit strukturiert planen, da alle Kinder ähnlich lange an einem Test arbeiten. Ebenso ist bei Computertests ein begrenzendes Zeitlimit notwendig, damit die Testung automatisch beendet wird. Solche Testungen können leicht im Unterricht und der Freiarbeitsphase per Tablet integriert werden.

Da für den Einsatz und die Anwendung im pädagogischen Alltag die Ökonomie der Tests für die Lernverlaufsdiagnostik ausschlaggebend ist, stellt die Bearbeitungsgeschwindigkeit der Schülerinnen und Schüler auch für die Testkonstruktion ein relevantes Merkmal dar. Der Fokus liegt dabei auf der Bearbeitungsgeschwindigkeit pro Item, da die gesamte Testzeit aufgrund des Zeitlimits festgesetzt ist. Durch die Tests sollen insbesondere Schülerinnen und Schüler mit Lernschwierigkeiten reliabel erfasst werden. Daher muss sichergestellt werden, dass auch in einem kurzen Test langsame Schülerinnen und Schüler ausreichend Items für eine aussagekräftige Auswertung des Speedtests bearbeiten können (Schurig, Jungjohann & Gebhardt, 2021). Je mehr Items pro Kind bearbeitet werden, desto mehr Informationen erhält die Lehrkraft. Gleichzeitig wird der Standardfehler des Testergebnisses geringer und das Ergebnis somit präziser.

Im Alltag nutzen Lehrkräfte die Ergebnisse eines schulischen Tests als Kriterium für das Erreichen eines Kompetenzziels. Die Festlegung von solchen Richtwerten ist bei Lernverlaufsdiagnostik und insbesondere bei Speedtests komplex.

Zum einen setzt sich das Ergebnis eines Speedtests aus der Anzahl an richtig bearbeiteten Aufgaben zusammen. Diese sind bei einer festgelegten Testzeit aber nicht nur von der Lösungswahrscheinlichkeit der Schülerin bzw. des Schülers, sondern auch von ihrer bzw. seiner Bearbeitungsgeschwindigkeit abhängig: Bei einer höheren Bearbeitungsgeschwindigkeit können mehr Aufgaben bearbeitet werden als mit einer langsameren Bearbeitungsgeschwindigkeit. Zusätzlich gibt es z. B. im Bereich Lesen verschiedene Arten der Bearbeitung. Es scheint sowohl schnelle Leserinnen und Leser zu geben, die viele Fehler machen, als auch langsame Leserinnen und Leser, die wenige Fehler machen. Beide Lesertypen erreichen eine ähnliche Anzahl an richtig bearbeiteten Aufgaben, aber eine unterschiedliche Bearbeitungsgeschwindigkeit. Die Bearbeitungsgeschwindigkeit wird meistens nicht separat betrachtet und ausgewertet, sondern durch die Festsetzung der Testzeit implizit angenommen und ausgedrückt.

Zum anderen ist das Vorgehen zur Ermittlung von Richtwerten für Lernverlaufstests bisher nicht einheitlich. Im Fall der Lernverlaufsdiagnostik gibt es zwei Möglichkeiten zur Festlegung von Richtwerten: Beim Curriculum Sampling Ansatz werden repräsentative Items aus dem Lehrplan ausgewählt und zur Festlegung von Richtwerten genutzt. Beim Ansatz des Robusten Indikators wird das Kriterium eher implizit einbezogen (Fuchs, 2004). So ist es fraglich, welcher Wert bei einem eindimensionalen Kompetenzbereich ein Kriterium darstellt, um die Förderung als erfolgreich bewerten zu können.

Außerdem ist das Raschmodell für die Lernverlaufsdiagnostik als Konstruktionsmodell empfohlen (Wilbert, 2014; Gebhardt et al., 2021). Dieses gibt dimensional-intervallskalierte Werte aus und repräsentiert keine ordinalen Schwierigkeitsstufen. Deswegen gibt es bei Tests, die auf dem Raschmodell aufbauen, keine natürlichen Cut-Off-Werte für die Bildung von Richtwerten.

Viele Faktoren, wie die Bearbeitungsgeschwindigkeit, die Testoberfläche oder die einbezogene Stichprobe können daher Richtwerte für Lernverlaufstests beeinflussen. Insbesondere die Bearbeitungsgeschwindigkeit ist je nach Erhebungsart oder -form, Situation und Test unterschiedlich (Bodmann & Robinson, 2004; Blumenthal & Blumenthal, 2020; Wirtz & Caspar, 2002). Richtwerte beziehen sich daher immer spezifisch auf einen Test und sind nur für diesen gültig.

Die Relevanz der Bearbeitungsgeschwindigkeit für die Testkonstruktion

Forschungsfragen

Eine Möglichkeit zur Herleitung von Richtwerten bei Speedtests stellen Hasbrouck und Tindal (1992; 2006; 2017) vor. Sie betrachten die Leseflüssigkeit gemessen in Wörtern pro Minute anhand eines unbekanntes Fließtextes. Pro Klassenstufe gibt es drei Messzeitpunkte, für welche die Perzentile 10, 25, 50, 75 und 90 dargestellt werden. Als Richtwerte empfehlen Hasbrouck und Tindal (2017) Rohwerte, welche bei Risikoschülern über dem Wert des Perzentils 10 (Hochrisiko) und des Perzentils 25 (Risiko) sowie bei regulären Schülerinnen und Schülern im Bereich plus oder minus 10 Wörter (Rohwerte) um das Perzentil 50 liegen. Ähnliche Vorgehen finden sich auch bei deutschen Testautorinnen und Testautoren (z. B. Jungjohann, Schurig & Gebhardt, 2021; Walter, 2014).

Die Betrachtung der Bearbeitungsgeschwindigkeit in Speedtests verspricht eine genauere Einschätzung der Ergebnisse von Schülerinnen und Schülern mit Lernschwierigkeiten und unterstützt die Ökonomie der Testkonstruktion. Somit stellen sich für Speedtests folgende Fragen:

- Fragestellung 1: Bearbeiten Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf Lernverlaufstests im Bereich Lesen langsamer als Schülerinnen und Schüler ohne sonderpädagogischen Unterstützungsbedarf?
- Fragestellung 2: Beeinflusst das Testalter die Bearbeitungsgeschwindigkeit?
- Fragestellung 3: Wie entwickelt sich die Bearbeitungsgeschwindigkeit im Testverlauf?
- Fragestellung 4: Ist die Bearbeitungsgeschwindigkeit geeignet, um kriteriale Richtwerte abzuleiten?

Methode

Zur Beantwortung der Fragestellung werden die Daten der Lesetests der Online-Plattform Levumi (www.levumi.de) ausgewertet. Levumi wird für die Forschung weiterentwickelt und bietet webbasierte digitale Tests der Lernverlaufsdagnostik sowie deren automatisierte Auswertung in den Bereichen Deutsch und Mathematik für Lehrende und Forschende frei zur Verfügung an (Jungjohann et al., 2018).

Instrumente

Im Lernbereich Deutsch sind Lesetests zur Messung der Leseflüssigkeit und sinnkonstruierendes Satzlesen implementiert, welche nach dem Prinzip des Robusten Indikators konstruiert, nach der Item-Response-Theory skaliert und in der dritten und vierten Jahrgangsstufe pilotiert sind (Jungjohann et al., 2021). In Anlehnung an verschiedene Lesemodelle gibt es verschiedene Tests zur Leseflüssigkeit und zum Leseverständnis: Silben lesen (SiL), Wörter lesen (WoL), Pseudowörter lesen (PWol), Buchstaben lesen (BuL) und Sinnentnehmendes Lesen (SinnL). Um auch in leistungsheterogenen Lerngruppen alle Schülerinnen und Schüler differenziert testen zu können, gibt es jeden Test mit bis zu fünf Niveaustufen. Die Tests sind als Speedtests angelegt. Im Bereich der Leseflüssigkeit dauern die Tests 60 Sekunden. Sie sind lehrerzentriert angelegt, das heißt, die Lehrpersonen überprüfen während des Vorlesens der Items deren Richtigkeit. Im Bereich des Leseverständnisses dauern die Tests fünf Minuten und die Schülerinnen und Schüler führen den Test selbstständig mit dem Tablet durch (Schurig et al., 2021). Während der Testung erfasst die Plattform die Antworten der Schülerinnen und Schüler als richtig oder falsch und speichert diese Informationen ab. Weiter wird die genaue Bearbeitungszeit pro Item in Millisekunden erfasst.

In der Pilotstudie von Jungjohann und Kollegen (2021) wurden für einige dieser Tests klassenstufenspezifisch querschnittliche Normen und längsschnittliche Zuwachsraten in Anlehnung an das Vorgehen von Walter (2011) und Shin und Forscherteam (2000) abgeleitet. Bei diesem Vorgehen wurden zuerst die Reliabilität basierend auf der Item-Response-Theory (IRT) geschätzt und anschließend anhand der Testrohwerte Richtwerte und Zuwachsraten für die .10-, .25-, .50-, .75- und 1.00-Perzentilgruppe separat pro Klassenstufen analysiert. Eine Berücksichtigung der Bearbeitungsgeschwindigkeiten hat dabei nicht stattgefunden. Die teilnehmenden Schülerinnen und Schüler (N = 90) besuchten die dritte und vierte Klassenstufe und waren zu Beginn der Studie 9;2 Jahre (SD = 0.69) alt. Zu Schuljahrsbeginn lösten die Kinder der dritten Klassenstufe im Test Wörterlesen 27.30 (SD = 12.09) und im Test Satzlesen 27.11 (SD = 13.56) Aufgaben korrekt. Die Schülerinnen und Schüler der vierten Klasse lösten in allen Tests mehr Aufgaben korrekt als die der dritten Klasse (Wörterlesen: 40.67 (SD = 11.80); Satzlesen: 39.00 (SD = 14.60)).

Dataming und ausgewählte Stichproben

Es werden die Erhebungen durch die fünf am häufigsten verwendeten Tests der Plattform (SinnL-N2, SinnL-N4, SinnL-N6, WoL-N4 und PWoL-N2a) ausgewertet. Der Datensatz wird in Form von Dataming aus den Testdaten der Plattform Levumi.de generiert. Bei den Daten handelt es sich einerseits um die erhobenen Daten von Lehrkräften aus der schulischen Praxis und andererseits um Daten von Forschenden, welche ihre Daten kontrolliert in ganzen Klassen erhoben haben. Der jeweilige Anteil pro Test ist:

- SinnL-N2: 88 % Lehrkräfte, 11 % Forschende
- SinnL-N4: 86 % Lehrkräfte, 14 % Forschende
- SinnL-N6: 16 % Lehrkräfte, 84 % Forschende
- WoL-N4: 79 % Lehrkräfte, 21 % Forschende
- PWoL-N2a: 98 % Lehrkräfte, 2 % Forschende

Für alle Tests SinnL-N2, SinnL-N4, SinnL-N6, WoL-N4 und PWoL-N2a zeigen die Ergebnisse der T-Tests, dass sich die erhobenen Daten durch Lehrkräfte und Forschende nicht signifikant voneinander unterscheiden und daher vergleichbar sind. Daher werden die Daten gesammelt ausgewertet.

Vorab werden die Rohdaten folgendermaßen bereinigt: Zum einen wird lediglich der jeweils erste Messzeitpunkt jedes Kindes in die Analyse miteinbezogen, da hier die Datengrundlage am stabilsten ist. Zum ersten Messzeitpunkt bearbeiten alle Schülerinnen und Schüler dieselbe vorab sortierte feste Reihenfolge, was die Vergleichbarkeit der Datensätze sicherstellt. Zum anderen werden Testdurchführungen mit einem Fehleranteil > 50% aussortiert, da hier von einer hohen Ratewahrscheinlichkeit ausgegangen werden kann. Tabelle 1 zeigt die Anzahl der Testdurchführungen vor und nach den Filterdurchgängen. Tabelle 2 zeigt die jeweilige Stichprobengröße sowie deren Verteilung in Hinblick auf das Geschlecht, das Alter und den Unterstützungsbedarf der Schülerinnen und Schüler. Die weiteren Auswertungen erfolgen mittels deskriptiver Statistik. Durch die digitale Umsetzung der Levumi-Tests kann die exakte Bearbeitungszeit pro Item in Millisekunden gemessen werden.

Test	A	B	C
SinnL-N6	2088	1224	1087
WoL-N4	1264	406	355
SinnL-N2	824	407	348
SinnL-N4	747	291	249
PWoL-N2a	443	188	160

Tabelle 1: Anzahl der Testdurchführungen in den Originaldaten (A), nach dem Filtern des ersten Messzeitpunkts (B) und nach hoher Ratewahrscheinlichkeit (C).

Tabelle 2: Ausgewertete Tests nach Geschlecht, Anzahl der Schülerinnen und Schüler mit und ohne sonderpädagogischen Unterstützungsbedarf und Alter

Test	Gesamt	Geschlecht				Sonderpädagogischer Unterstützungsbedarf				Alter M (SD)			
		w	m	d	k. A.	k. SUB	SUB L.	SUB GE.	sonstige	k. A.	gesamt	SUB. L.	SUB. GE.
SinnL-N6	1089	456 (42%)	601 (55%)	0 (0%)	32 (3%)	591 (54%)	161 (15%)	11 (1%)	197 (18%)	129 (12%)	12,85 (2,20)	13,07 (2,32)	10,21 (3,04)
WoL-N4	355	115 (32%)	127 (36%)	1 (0,3%)	112 (32%)	24 (7%)	29 (8%)	10 (3%)	25 (7%)	267 (75%)	9,15 (2,63)	14,05 (4,87)	13,88 (0,9)
SinnL-N2	349	113 (32%)	174 (50%)	2 (0,5%)	60 (17%)	48 (14%)	77 (22%)	16 (5%)	42 (12%)	166 (48%)	9,55 (2,09)	10,6 (1,95)	12,3 (2,22)
SinnL-N4	251	103 (41%)	129 (51%)	1 (0,4%)	18 (7%)	60 (24%)	44 (18%)	5 (2%)	18 (7%)	124 (49%)	10,35 (1,42)	11,21 (1,16)	12 (3,31)
PWoL-N2a	160	74 (46%)	83 (52%)	1 (0,6%)	2 (1%)	5 (3%)	69 (41%)	3 (2%)	8 (5%)	75 (47%)	8,92 (1,4)	9,63 (1,8)	k.A.

Ausgewertete Tests unter Berücksichtigung von Größe der Gesamtstichprobe (Gesamt), Geschlecht (Geschlecht), Schülerinnen und Schüler ohne sonderpädagogischen Unterstützungsbedarf (k. SUB), Unterstützungsbedarf im Lernen (SUB L), Unterstützungsbedarf in der geistigen Entwicklung (SUB G), sonstigen Unterstützungsbedarfen (sonstige) und Schülerinnen und Schüler ohne Angabe eines Unterstützungsbedarf (k. A.), mittleres Alter der Kinder (Alter)

Zur Beantwortung der ersten Forschungsfrage wurde die durchschnittliche Bearbeitungszeit pro Item für jeden Schüler und jeden Test berechnet (siehe Tabelle 3, Abbildung 1). Diese Ergebnisse wurden innerhalb einer Gruppe mit jeweiligem sonderpädagogischen Bedarf an Unterstützung erneut gemittelt, um die durchschnittliche Bearbeitungszeit pro Item eines Schülers in der jeweiligen Gruppe zu erhalten. Dabei zeigt sich, dass Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf die Items in vier von fünf Tests langsamer bearbeiten als Schülerinnen und Schüler ohne Unterstützungsbedarf ($F(7,759) = 25,867, p < .001$). Außer-

Ergebnisse



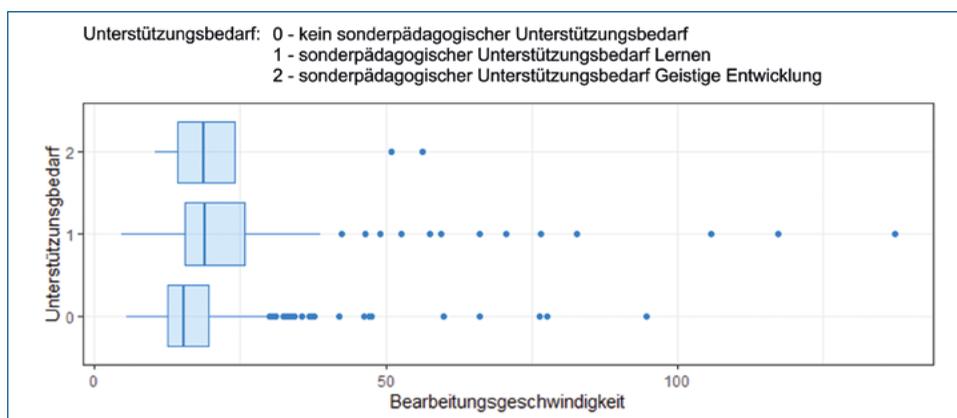
dem ist die Heterogenität der Bearbeitungsgeschwindigkeiten bei Schülerinnen und Schülern mit sonderpädagogischem Unterstützungsbedarf höher (Kein Unterstützungsbedarf: $s^2 = 68186431$, Unterstützungsbedarf: $s^2 = 319235777$). Die Variabilität wird beispielhaft für den Test SinnL-N6 visualisiert (Abbildung 1), da für diesen Test die größte Datengrundlage vorliegt. An den eingetragenen Mittelwerten der Boxplots ist erkennbar, dass Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf im Lernen oder der Geistigen Entwicklung durchschnittlich langsamer arbeiten und es eine größere Breite der Bearbeitungszeiten gibt.

*Tabelle 3:
Durchschnittliche
Bearbeitungsgeschwindigkeit
pro Item*

Test	Durchschnittliche Bearbeitungsgeschwindigkeit pro Item M (SD)					
	Kein SUB		SUB Lernen		SUB Geistige Entwicklung	
SinnL-N6	17.17	(8.26)	24.16	(18.07)	23.8	(15.42)
WoL-N4	2.6	(0.9)	2.4	(0.8)	2.9	(1.6)
SinnL-N2	17.34	(15.14)	19.28	(16.04)	24.7	(24.17)
SinnL-N4	12.8	(6.3)	15.8	(8.5)	24.3	(11.2)
PWoL-N2a	3.9	(1.8)	4.3	(2.4)	3.4	(0.4)

Mittlere Bearbeitungsgeschwindigkeit in Sekunden pro Item für gesamte Schülerschaft getrennt nach sonderpädagogischem Unterstützungsbedarf: Kein sonderpädagogischer Unterstützungsbedarf (Kein SUB), sonderpädagogischer Unterstützungsbedarf im Lernen (SUB Lernen), und Unterstützungsbedarf in der geistigen Entwicklung (SUB Geistige Entwicklung)

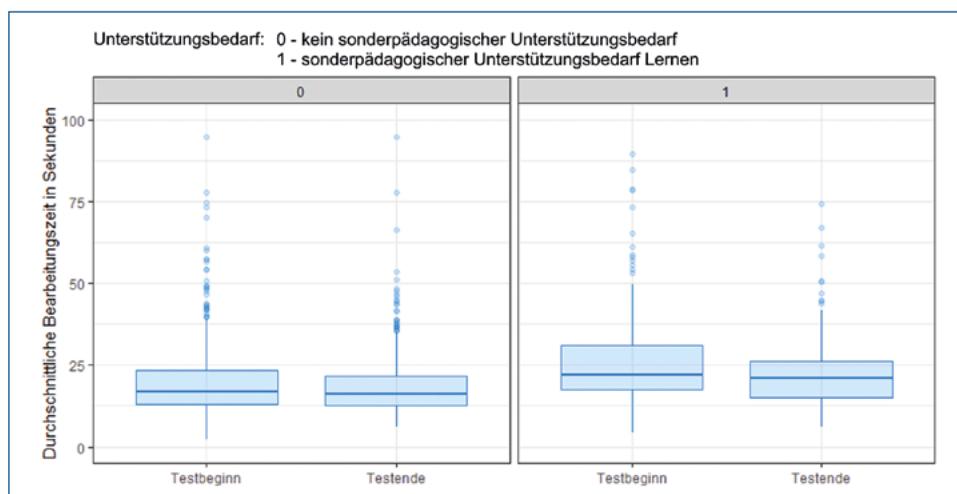
*Abbildung 1:
Durchschnittliche
Bearbeitungsgeschwindigkeit
für den Test SinnL-N6*



Zur Beantwortung der zweiten Forschungsfrage wurde das Alter der Schülerinnen und Schüler zum Testzeitpunkt in Relation zu ihrer durchschnittlichen Bearbeitungszeit im Test SinnL-N6 gesetzt. 23% der Schülerdaten konnten aufgrund fehlender Angaben zum Alter nicht in diese Analyse einbezogen werden. Die durchschnittliche Bearbeitungszeit pro Item korreliert kaum mit dem Alter der Schülerinnen und Schüler ($r[850] = -0.178, p = .000$). Die Bearbeitungszeit ist in Bezug zum Alter breit gestreut.

Zur Beantwortung der dritten Forschungsfrage wurde die Geschwindigkeit der Schülerinnen und Schüler während eines Testdurchgangs des SinnL-N6 betrachtet. Durch die genaue Messung der Bearbeitungszeit pro Item in Millisekunden lässt sich die Geschwindigkeit pro Item jedes Schü-

*Abbildung 2:
Durchschnittliche
Bearbeitungsgeschwindigkeit pro Item
am Beginn und Ende des Tests
SinnL-6 – Trennung nach sonderpädagogischem Unterstützungsbedarf
Lernen*



lers nachvollziehen. Die Geschwindigkeit am Testbeginn wurde aus den ersten drei bearbeiteten Items gemittelt, die Geschwindigkeit am Testende wurde aus den letzten drei bearbeiteten Items gemittelt (siehe Abbildung 2). Die Unterschiede zwischen der Bearbeitungszeit zu Beginn und zum Ende des Tests sind nicht signifikant, das heißt die Items werden am Anfang und am Ende vergleichbar schnell bearbeitet.

Test	Summenscore M (SD)			Fehlerscore M (SD)			Fehleranteil in %		
	k. SUB	SUB L	SUB GE	k. SUB	SUB L	SUB GE	k. SUB	SUB L	SUB GE
SinnL-6	18.9 (7.0)	14.1 (6.2)	15.4 (7.6)	2.4 (1.7)	3.8 (7.9)	2.1 (1.6)	12.2	18.1	11.4
WoL-4	22.4 (7.2)	25.2 (9.2)	21.4 (10)	3.5 (1.8)	3.7 (2.09)	4.8 (4.3)	13.6	12.9	18.3
SinnL-2	23.4 (13.99)	23.08 (15.4)	16.68 (11.21)	3.46 (2.89)	4.27 (3.53)	3.92 (3.23)	12.9	15.6	19
SinnL-4	25.8 (11)	19.1 (9)	14.4 (5.2)	3.2 (3)	5.4 (3.4)	3.5 (2.1)	11	22	19.6
PWoL-2a	15.6 (5.3)	15.8 (12)	13 (5.3)	2.2 (1.1)	3.8 (2.1)	5.3 (3.8)	12.4	19.5	29

Trennung nach sonderpädagogischem Unterstützungsbedarf: Kein Unterstützungsbedarf (k. SUB), sonderpädagogischer Unterstützungsbedarf im Lernen (SUB L), und sonderpädagogischer Unterstützungsbedarf in der geistigen Entwicklung (SUB GE).

Zur Beantwortung der letzten Forschungsfrage wurde zunächst der durchschnittliche Anteil an richtig und falsch gelösten Items der nach Unterstützungsbedarf gruppierten Schülerinnen und Schüler ermittelt und daraus der Fehleranteil in % abgeleitet (siehe Tabelle 4). In drei von fünf Tests bearbeiten Schülerinnen und Schüler ohne Unterstützungsbedarf mehr Items richtig und in allen Tests weniger Items fehlerhaft. Der Fehleranteil ist dementsprechend in vier von fünf Tests geringer. Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf im Lernen weisen eine höhere Standardabweichung in drei von fünf Tests in den Summenscores und in allen Tests in den Fehlerscores auf. Die Verteilung von Summenscore, Fehlerscore und Fehleranteil bei Schülerinnen und Schülern mit sonderpädagogischem Unterstützungsbedarf in der geistigen Entwicklung ist heterogen und nicht systematisch.

Tabelle 4:
Summenscore, Fehlerscore und Fehleranteil in % und resultierender Richtwert für die einzelnen Tests

Auf dieser Grundlage lassen sich verschiedene Richtwerte für die Tests annehmen. In Bezug auf die Ergebnisse von Hasbrouck und Tindal (2017) werden hierfür die Perzentile 10, 25, 50, 75 und 90 der mittleren Summenscores aller Schülerinnen und Schüler als mögliche Richtwerte berechnet (siehe Tabelle 5).

Test	.10	.25	50	.75	90
SinnL-6	9	13	18	22	27
WoL-4	12	15	22	29	37
SinnL-2	6	11	18	27	43
SinnL-4	10	15	22	31	40
PWoL-2a	7	10	16.5	24	29

Tabelle 5:
Mögliche Richtwerte für die Anzahl richtig zu lösender Items mit Basis der Perzentile 10, 25, 50, 75 und 90

Lernverlaufsdiagnostik sollte so kurz wie möglich sein, um ökonomisch im Klassenzimmer umsetzbar zu sein und die Konzentration der Schülerinnen und Schüler nicht zu überfordern. Trotzdem sollten Schülerinnen und Schüler in der vorgegebenen Zeit genug Items bearbeiten können, damit auch bei Speedtests ein reliables Ergebnis des aktuellen Standes gemessen wird. Um solche Tests zu entwerfen, benötigt man Informationen über die Bearbeitungsgeschwindigkeit des Kindes sowie Informationen über die ungefähre Entwicklungsstufe der gemessenen Kompetenz. Die Bearbeitungsgeschwindigkeit ist aber je nach Erhebungsart, -form und Test unterschiedlich (Bodmann & Robinson, 2004; Blumenthal & Blumenthal, 2020). Aus diesem Grund wurden in diesem Beitrag die Daten der Testplattform Levumi überprüft. Die Plattform erfasst Daten von Forschungsstudien und Daten aus der Praxis, die in beiden Fällen in die Analyse mit einfließen. Das bedeutet, dass sowohl in einer standardisierten Testsituation mit geschultem Personal, als auch im Unterrichtsalltag Daten erhoben wurden, was die Bearbeitungsgeschwindigkeit der Kinder beeinflussen könnte (Wirtz & Caspar, 2002). In Bezug auf die Bearbeitungszeit konnte hierbei kein Unterschied zwischen den Daten aus der Praxis und den Daten aus den Forschungsstudien in den durch Datamining generierten Daten festgestellt werden.

Diskussion

Eine Limitation dieser Daten ist, dass die Daten der Forschenden meist alle zusätzlichen Schülerinformationen (Alter, Geschlecht und Unterstützungsbedarf) enthalten. Die Daten aus der schuli-



Schlüsselwörter

Lernverlaufsdiagnostik, Kriterienorientierung, Lesen, Individualisierter Unterricht, Leistungsrückmeldungen

Abstract

Progress monitoring tests must be as short as possible in order to be used economically in the classroom. To ensure reliable measurement, it is necessary that students with learning difficulties are able to complete an adequate number of items during the test period. This article describes an approach based on the average processing speed per item on the online platform Levumi in reading. Students with special needs complete the speed-tests slower and make more mistakes despite a higher average age at the time of testing. The conclusions that can be drawn on the basis of the processing speed and the extent to which guidelines for defining a good test result can be derived are discussed.

Keywords

Learning progress diagnostics, criteria orientation, reading, individualized teaching, performance feedback

Praxis sind in diesem Bereich jedoch nicht vollständig. So wurde im Test WoL-4 beispielsweise zu 75 % der Praxisdaten nicht erfasst, ob ein sonderpädagogischer Unterstützungsbedarf vorliegt. Daraus resultieren Datenlücken, aufgrund derer nicht alle Testdaten exakt bezüglich des Unterstützungsbedarfs differenzierbar sind. Es ist daher notwendig, sowohl Forscher- als auch Praxisdaten in der Analyse heranzuziehen. Die Daten aus der Praxis sind zwar ökonomischer, jedoch müssen sie kritischer betrachtet und mehrfach geprüft werden.

Die Auswertungen zeigen, dass Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf in Lesetests durchschnittlich langsamer und fehlerhafter arbeiten (Fragestellung 1). Daher müssen die Tests so konzipiert werden, dass sie im unteren Bereich differenzieren und faire Ergebnisse liefern. Selbiges wurde für Levumi bereits getestet (Jungjohann, DeVries, Mühling & Gebhardt, 2018). Lehrkräfte wählen bei der Arbeit mit Levumi die passenden Tests selbst aus. Daraus resultiert, dass Schülerinnen und Schüler mit Lernschwierigkeiten durchschnittlich ein Jahr älter bei der Testbearbeitung sind als andere Schülerinnen und Schüler bei dem gleichen Test. Daher ist das Alter nicht ausschlaggebend für die Bearbeitungsgeschwindigkeit der Schülerinnen und Schüler in den Levumi-Speedtests (Fragestellung 2). Kinder mit Lernschwierigkeiten erreichen später das geeignete Lesekompetenzniveau, mit welchem die Bearbeitung der jeweiligen Tests vorgesehen ist (Klicpera et al., 1993).

Schülerinnen und Schüler mit Lernschwierigkeiten weisen ein Defizit im Arbeitsgedächtnis auf, welches in engem Zusammenhang mit der kognitiven Geschwindigkeit und der Aufmerksamkeit steht (Tourva et al., 2016; Schuchardt et al., 2010). In Bezug auf die Testsituation ergaben sich insgesamt Unterschiede in der Geschwindigkeit. Die Bearbeitung der Tests erfolgt aber pro Kind mit konstanter Geschwindigkeit. Es ist nicht der Fall, dass Schülerinnen und Schüler mit Unterstützungsbedarf am Ende der Testzeit langsamer werden. Eine langsamere Bearbeitung am Ende wäre ein Indiz

für zunehmende Konzentrationsschwierigkeiten. Die Bearbeitungsgeschwindigkeit der Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf blieb während der Tests konstant. Somit ist anzunehmen, dass die Konzentrationsfähigkeit der Kinder und Jugendlichen hinreichend ist, um die Leistung während der kurzen Dauer der Tests von einer bzw. fünf Minuten konstant aufrecht zu erhalten.

Richtwerte helfen der Praxis bei der Einschätzung von Testergebnissen. Generell sollten Schülerinnen und Schüler nur 5 % Fehler beim Lesen machen. Die durchschnittlichen Ergebnisse zeigen aber Fehlerraten zwischen 10 und 15 %. Ein weiterer Indikator für Richtwerte ist der Summenscore (Fragestellung 5). Hasbrouck und Tindal (1992; 2006; 2017) empfehlen dabei die Orientierung an niedrigen Perzentilen für schwache und an hohen Perzentilen für starke Schülerinnen und Schüler. Verschiedene Richtwerte für unterschiedliche Personengruppen oder Unterstützungsbedarfe erscheinen nach den Ergebnissen nicht als sinnvoll, da die Bearbeitungsgeschwindigkeiten ähnlich sind und die Lehrkraft auch für Kinder mit einem sonderpädagogischen Unterstützungsbedarf die passenden Tests wählt. Somit messen die Tests ein bestimmtes Konstrukt reliabel, wenn die Zielgruppe passend ausgewählt wurde (Anderson et al., 2020). Deswegen ist anzunehmen, dass Schülerinnen und Schüler mit Unterstützungsbedarf im Hinblick auf eine individuelle Bezugsnorm die gleichen Ergebnisse erreichen können wie andere Schülerinnen und Schüler. Sie benötigen lediglich eine längere Übungszeit. Eigene Richtwerte für schwache Schülerinnen und Schüler sind nur dann notwendig, wenn der reguläre Richtwert nie erreicht werden kann. Dies kann auch der Fall sein, weswegen Levumi auf die Testauswahl durch die Expertise der Lehrkraft setzt und keine feststehenden Klassenstufentests für alle Schülerinnen und Schüler anbietet. Die Frage nach objektiven Richtwerten ist daher keine empirische, sondern eine normative Frage.

Für die Praxis lassen sich hieraus mehrere Aspekte ableiten. Schülerinnen und Schüler mit einem sonderpädagogischen Unterstützungsbedarf im Lernen oder der geistigen Entwicklung arbeiten

langsamer und machen mehr Fehler als Schülerinnen und Schüler ohne Unterstützungsbedarf. Es empfiehlt sich daher, insbesondere diese Schülergruppe weiter in ihrem Lernprozess intensiv zu begleiten, um wertvolle Übungs- und Lernzeit mit effektiven Förderungen gestalten zu können. Lernverlaufsdiagnostik stellt hier eine geeignete Möglichkeit dar. Dabei kann nicht nur die Anzahl an richtig gelösten Antworten ausgewertet, sondern zusätzlich die Bearbeitungsgeschwindigkeit anhand der Anzahl der insgesamt bearbeiteten Aufgaben betrachtet werden.

Digitale Lernverlaufsdiagnostik ist dabei praktisch anzuwenden und bietet weitere Vorteile für die Schulpraxis. Die Ergebnisse der Lernverlaufstests werden automatisch ausgewertet und direkt an die Lehrperson zurückgemeldet. Die digitale Umsetzung bietet zudem die Möglichkeit, die Lesegeschwindigkeit sensibel zu erfassen und den individuellen Unterstützungsbedarf im Lernen dadurch genau zu analysieren.

Auf das beobachtete Bearbeitungsverhalten von Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf können Lehrpersonen in beispielsweise anschließenden Feedbackgesprächen sensibel reagieren. Die Lehrkräfte können den Kinder und Jugendlichen ihr Aufrechterhalten der Konzentration und Aufmerksamkeit über die Testdauer hinweg positiv anmerken. Das kann das Selbstwertgefühl und die Motivation nicht nur für folgende Testungen, sondern auch in schulischen Lernsituationen erhöhen.

Die Bearbeitungszeit ist auch im Unterrichtsalltag, in Einzelarbeitsphasen und bei der Planung und Durchführung von selbst entwickelten Verfahren von Schülerinnen und Schülern ein wichtiger Faktor. Sie sollte daher dokumentiert und für die Förderplanung ausgewertet werden. Da die Bearbeitungsgeschwindigkeit innerhalb von fünf Minuten durchgehender Arbeitszeit nicht langsamer wird, leiten wir weiter daraus ab, dass kurze und intensive Arbeitsphasen von fünf Minuten Schülerinnen und Schüler mit Unterstützungsbedarf Lernen hinsichtlich ihres Konzentrationsvermögens nicht benachteiligen. Sie können daher in Förderphasen integriert werden können, ohne die Fairness für Schülerinnen und Schüler ohne Unterstützungsbedarf zu schwächen. Um die Leselernprozesse bei Schülerinnen und Schülern mit sonderpädagogischem Unterstützungsbedarf im Lernen zu unterstützen, kann somit in mehreren kurzen Arbeitsphasen statt einer langen Arbeitsphase geübt werden. In kurzen Arbeitsphasen können auch Schülerinnen und Schülern mit sonderpädagogischem Unterstützungsbedarf im Lernen ihre Konzentration aufrecht halten und die verfügbare Zeit mit Lernaktivitäten füllen.

Die für den Beitrag verwendete Literatur steht Ihnen online zur Verfügung:

www.verband-sonderpaedagogik.de/wp-content/uploads/2022/12/zfh_1.2023_literatur-ebenbeck.pdf

Der Beitrag mit allen verwendeten Quellen findet sich ebenfalls in der digitalen Bibliographie.



Literatur

Nikola Ebenbeck
Lehrstuhl für Lernbehindertenpädagogik
einschließlich inklusiver Pädagogik
Universität Regensburg
Sedanstraße 1
93055 Regensburg
nikola.ebenbeck@paedagogik.uni-regensburg.de

Andreas Mühling
Arbeitsgruppe Didaktik der Informatik
Christian-Albrechts-Universität zu Kiel
Christian-Albrechts-Platz 4
24118 Kiel

Jana Jungjohann
Lehrstuhl für Lernbehindertenpädagogik
einschließlich inklusiver Pädagogik
Universität Regensburg
Sedanstraße 1
93055 Regensburg

Markus Gebhardt
Lehrstuhl für Lernbehindertenpädagogik
einschließlich inklusiver Pädagogik
Universität Regensburg
Sedanstraße 1
93055 Regensburg