LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Pädagogische Diagnostik

Leistung, Kompetenz und Entwicklung
messen, bewerten und für individuelle Förderung interpretieren

Markus Gebhardt
Professor für Sonderpädagogik – Förderschwerpunkt Lernen einschließlich inklusiver Pädagogik
Ludwig-Maximilians-Universität München
https://orcid.org/0000-0002-9122-0556

Version 0.3
März 2023

Veröffentlicht unter der Lizenz:
CC-BY-NC-SA 4.0

Münchner Beiträge zur Inklusions- und Sonderpädagogik ; Band 3

Münchner Beiträge zur Inklusions- und Sonderpädagogik ; Band 3
herausgegeben von Markus Gebhardt

ISSN 2942-7193

Zusammenfassung

Pädagogische Diagnostik ist eine Erkenntnisbemühung für die Grundlage von pädagogischen Entscheidungen. Im Alltag von Bildungseinrichtungen werden Leistungen und Kompetenzen implizit oder explizit bewertet. Immer dann findet auch pädagogische Diagnostik statt. Pädagogische Diagnostik ist eine nachvollziehbare und dokumentierbare Handlung mit dem Ziel, pädagogische Maßnahmen auszuwählen, zu begleiten und zu beobachten. Dieses Buch richtet sich an alle Personen in Bildungseinrichtungen und insbesondere an Studierende des Lehramts, der Sonderpädagogik und Schulpsychologie. Das Buch ist als Einführungswerk angelegt und stellt die Grundbegriffe und Konzepte von pädagogischer Diagnostik in Bildungseinrichtungen vor. Hierbei werden die verschiedenen Formen von Tests sowie die Grundlagen des Messens skizziert und in Bezug auf die pädagogische Diagnostik von Lernschwierigkeiten dargestellt.

Schlagwörter

Diagnostik, Inklusion, Förderschule, Lernschwierigkeiten, Förderbedarf, Unterstützungsbedarf, sonderpädagogischer Schwerpunkt, Lernen, Response to Intervention, Lernverlaufsdiagnostik

Zitierung:

Gebhardt, M. (2023). *Pädagogische Diagnostik.* Leistung, Kompetenz und Entwicklung messen, bewerten und interpretieren für individuelle Förderung. (Version 0.3). Ludwig-Maximilians-Universität München.

Inhaltsverzeichnis

1 Diagnostik als präventive Grundlage für pädagogische Handlungsentscheidung 1
1.1 Diagnostik als Teil der inklusiven Schule 3
1.2 Pädagogische Diagnostik 7
1.3 Didaktik und Diagnostik 11
1.4 Der diagnostische Prozess 12
1.5 Von der Person zu den Umweltbedingungen 14
1.6 Ebenen der Diagnostik 17
1.6.1 Diagnostik auf der Ebene der Person 17
1.6.2 Diagnostik auf der Ebene des häuslichen und schulischen Umfelds 17
1.6.3 Das Einzelfallraster für die pädagogische Diagnostik 20
1.7 Glossar – Übersicht über Begriffe im Bereich Diagnostik 20
2 Statusdiagnostik oder Verlaufsdiagnostik 22
2.1 Modelle der Prozess- oder Förderdiagnostik 23
2.2 Modell zur Lernverlaufsdiagnostik 25
2.3 Zone der nächsten Entwicklung 28
2.4 Data Based Decision Making 29
3 Sonderpädagogische Diagnostik 31
3.1 Ziele der sonderpädagogischen Diagnostik 31
3.2 Erstellung eines sonderpädagogischen Gutachtens 34
3.4 Förderplan 36
4 Messen und Bewerten im pädagogischen Kontext 38
4.1 Lernen und die Entwicklung von schulischen Kompetenzen 40
4.2 Beobachtung zum Lernen und qualitative Fehleranalyse 43
4.3 Diktate und Proben als Prüfungsform 44
4.4 Latente Konstrukte und Dimensionen 45
4.5 Latente Dimensionen im Fragebogen 49
4.6 Normen 50
4.7 Grundlagen der Testtheorie 54
4.8 Gütekriterien 55
4.9 Messinvarianz und Adaptionen 58
4.10 Testeffekte 59
5 Überblick über verschiedene Formen von diagnostischen Instrumenten 60
5.1 Power- oder Speedtest 61
5.2 Klassenstufen- oder konstruktbasierte Schulleistungstests 62
5.3 Regelbasierte Itemkonstruktion 63
5.4 Kriterienorientierte Tests 64
5.5 Screenings 65

5.6 Verlaufsdiagnostiktests 66
5.7 Adaptive Tests 68
5.8 Lernspiele, Games und Übungen mit Rückmeldungen während des Tests 70
6 Leitfäden zur pädagogischen Diagnostik 72
6.1 Wie wählt man einen Test aus? 72
6.2 Wie wird eine faire Probe oder einen informellen Test konstruiert? 73
6.3 Auswertung von standardisierten Screenings 76
6.4 Auswertung von Lernverlaufsdiagnostik 77
7 Diagnostik und Förderung 78
7.1 Direkte Instruktion 78
7.2 Direkte Instruktion und/oder Problem Based Learning 81
7.3 Diagnostik und Förderung mit Levumi 83
7.4 Ablaufschema Praktikum Förderdiagnostik 84
8 Einzelfallforschung 85
9 Beratung und Elterngespräche 87
10 Ausblick 88
11 Anhang: Hinweise zu einzelnen Bereichen 89
11.1 Kognitive Voraussetzungen und Intelligenz 89
11.2 Schriftspracherwerb & Lesen 91
11.3 Rechtschreiben 92
11.4 Mathematik 93
11.5 Verhalten und Erleben 94
12 Literatur 96

Vorwort

Pädagogische Diagnostik ist ein wichtiger Baustein der Ausbildung in allen pädagogischen Berufen. Die Diagnostik ermöglicht es, einzelne Personen oder Gruppen einzuschätzen, pädagogische Handlungen und Maßnahmen auszuwählen und diese Auswahl zu begründen. Dieses Buch dient als Einführungswerk in die pädagogische Diagnostik. Hierbei richtet sich das Buch an alle interessierten Leser:innen aus allen pädagogischen Bereichen. Die Grundlagen und der Inhalt sind daher für alle pädagogischen Berufe relevant. Bei den Beispielen bezieht sich das Buch auf Lernschwierigkeiten und es wird ein sonderpädagogischer und pädagogisch-psychologischer Blick vertreten.
Dieses Buch baut auf den pädagogischen Grundlagen des Buches „Inklusiv- und sonderpädagogische Pädagogik im Schwerpunkt Lernen. Eine Einführung" auf und bezieht das Fallbuch zum sonderpädagogischen Schwerpunkt Lernen sowie das Herausgeberwerk „Handbuch sonderpädagogische Diagnostik" mit ein, welche ebenfalls kostenfrei als PDF Open Access veröffentlicht wurden. Zur Einführung in die Diagnostik gibt es bereits einige sehr gute Lehrbücher. Meist beinhalten diese die jeweilige Fachlichkeit der:s Autor:in. So richten sich die Bücher der Psycholog:innen an Psycholog:innen, während Fachdidaktiker:innen sich meist auf ihr schulisches Fach beziehen. Dieses Buch fokussiert sich auf die pädagogische Diagnostik in der inklusiven Schule und erklärt die Grundlagen der Testtheorie in Bezug auf die pädagogischen Handlungen, die gängigen Konzepte zur Statusdiagnostik und der Gutachtenerstellung als auch die Prozess- bzw. Verlaufsdiagnostik für eine anschließende Förderplanung.

Eine Herausforderung ist hierbei, dass pädagogischen Studiengängen, insbesondere Lehramtsstudiengängen, vielfach notwendige Vorlesungen zur Statistik und Testtheorie fehlen. Aus diesem Grund bezieht das Buch die wichtigsten Themen mit ein und verweist auf fachlich tiefergehende Werke. So gibt es zahlreiche Methodenbücher, welche von Studierenden des Lehramts und Lehrkräften eher selten gelesen werden, und gute pädagogische Bücher sowie didaktische Ratgeber, welche kaum auf die Testkonstruktion eingehen. Dieses Buch stellt sich dieser Herausforderung und versucht, diese Bereiche in zugänglicher Sprache darzustellen.

In diesem Buch wird hierbei kein maßgeblicher Unterschied zwischen pädagogischer und psychologischer Diagnostik angenommen. Beide diagnostischen Ansätze benötigen in den pädagogischen Handlungsfeldern die Erkenntnisse der empirischen Bildungsforschung und psychometrisch geprüfte Instrumente. Die spätere Berufspraxis und die Aufgabenfelder unterscheiden sich jedoch zwischen Psycholog:innen, Schulpsycholog:innen, Sonderpädagog:innen, Sozialpädagog:innen sowie Klassen- und Fachlehrkräften.

Dieses Buch ist in Open Access veröffentlicht und möchte auch bei seinen Quellen und Verlinkungen möglichst frei zugängliche Werke benutzen. Daher wurden Artikel hinter einer Bezahlschranke nur spärlich verlinkt oder zitiert. Das Buch wird in den nächsten Semestern anhand der Anregungen von Kolleg:innen und Student:innen verbessert und veröffentlicht.

Einsatz des Buches in Vorlesungen und Seminaren

Dieses Buch dient als hilfreiches Grundlagenwerk für mehreren Kursen an der LMU München zur Diagnostik und Didaktik wie dem Fallseminar, der Vorbereitung zum sonderpädagogischen Praktikum und zur Begleitung des Praktikums. Ziel dieser Kurse ist es ein grundlegendes Verständnis für diagnostische Prozesse in Schule und Bildungseinrichtung zu schaffen. Vorgaben zum Einsatz und der Gestaltung von Diagnostik sind in jedem Bundesland und teilweise in einzelnen Schulformen anders umgesetzt, daher sind im Buch keine Gesetzesvorgaben zitiert, sondern die verbindenden dahinterliegenden Konzepte diskutiert.

Studierende haben die Aufgabe im Verlauf ihres Studiums informelle Testverfahren, Gutachten, Förderpläne und Fördermaterial zu erstellen. Diese Aufgaben wurden als Prüfungsformen gewählt, da eine Klausur zu pädagogischer Diagnostik nur an der Oberfläche kratzen würde und die Frage zur Übertragung der Erkenntnisse durch die Studierenden in die Praxis bei einer Klausur offenbliebe. So ist eine Modulprüfung die Erstellung eines eigenen informelles Verfahrens. Studierende wählen einen Lernbereich aus und überlegen für welche Form der Diagnostik das Instrument einsetzt und verwendet werden könnte. In diesem Bereich vertiefen Studierende ihr Wissen und ein anwendungsorientierter Bezug wird erworben. Zur Erstellung eines informellen Verfahrens gibt es mehrere Arbeitsschritte aus, deren Hintergründe in diesem Buch erklärt werden:

- Themenbereich sowie zu messendes Konstrukt für ein informelles Verfahren auswählen,
- zu messende Kompetenzstufen oder Messbereich nach wissenschaftlichen Erkenntnissen und Konstrukten sowie gegebenenfalls nach dem Lehrplan bestimmen,
- bestehende Verfahren prüfen und vergleichen,
- eigene Zielsetzung und Anwendung des informellen Verfahrens definieren,
- Dimensionalität festlegen und Regeln der Item-Konstruktion überlegen,
- Item-Pools entwickeln,
- ein oder mehrere Testhefte aufbauen,
- Pilotierung - wenn möglich – durchführen und
- Ergebnisse reflektieren und im Portfolioansatz oder als Seminararbeit dokumentieren.

Übersicht über die Videos entlang der Ordnung der Vorlesung

Nr.	Titel	Dauer
01	Diagnostik in der Schule	(25:15)
02	Pädagogische Diagnostik	(08:25)
03	Bezugsnormenorientierung	(14:17)
04	Feststelldiagnostik, Förderdiagnostik und Lernverlaufsdiagnostik	(16:22)
05	Skalenniveaus	(10:50)
06	Median, Mittelwerte, Boxplot und Normalverteilung	(16:38)
07	Korrelationen	(10:21)
08	Manifeste und latente Konstrukte	(13:47)
09	Unabhängige und abhängige Variablen	(9:53)
10	Normen bei Tests	(17:48)
11	Lernverlaufsdiagnostik	(15:48)
12	Videoreihe von David Scheer zur Einzelfallforschung	

Generelle Buchempfehlungen:

Bundschuh, K. & Winkler, C. (2019). *Einführung in die Sonderpädagogische Diagnostik*. UTB.

Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion*. Pearson.

Gebhardt, M., Jungjohann, J. & Schurig, M. (2021). *Lernverlaufsdiagnostik im förderorientierten Unterricht. Eine Einführung in Anwendung, Testkonstruktion und Instrumente.* Reinhardt Verlag.

Gebhardt, M., Scheer, D. & Schurig, M. (2022). *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung.* Universität Regensburg.

Kubinger, K. D. (2009). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Hogrefe.

Tröster, H. (2019). *Diagnostik in schulischen Handlungsfeldern. Methoden, Konzepte, praktische Ansätze.* Kohlhammer.

Watkins, A. (2007). *Assessment in Inclusive Settings: Key Issues for Policy and Practice.* European Agency for Development in Special Needs Education.

1 Diagnostik als präventive Grundlage für pädagogische Handlungsentscheidung

Pädagogische Maßnahmen haben das Ziel, Personen zu nutzen und zu unterstützen. Hilft eine pädagogische Handlung nicht oder nicht ausreichend, sollte diese Handlung nicht weiterführt werden oder durch andere Handlungen ersetzen. Ziel pädagogischer Diagnostik ist es, möglichst gut anhand von nachvollziehbaren Schritten und dokumentierten Handlungen pädagogische Entscheidungen zu begründen, so dass diese auch für Außenstehende, aber besonders für die betreffende Person, für die Eltern und auch dem pädagogischen Team klar sind. Dieser Prozess wird in der amerikanischen Forschung als Data Based Decision Making bezeichnet, da Entscheidungen infolge von Beobachtungen, Ergebnissen von Fragebögen und Tests Daten zu einer Person entstehen, die gemeinsam im Team interpretiert werden. Hierbei wird in der amerikanischen Forschung nicht den Begriff Diagnostik, sondern eher Educational Assessment (Brookhart & McMillan, 2020) verwendet, um dies von medizinischer Diagnostik abzugrenzen.

Ziel pädagogischer Diagnostik ist es, möglichst präventiv, nachvollziehbar anhand von Standards, möglichst schnell und benutzerfreundlich pädagogische Fragen zu beantworten, um eine pädagogische Situation und Problemstellung dauerhaft und nachhaltig zum Positiven für alle Beteiligten zu verändern. Die pädagogische Fragestellung steht im Vordergrund. Pädagogische Diagnostik ist ein Bemühen um Erkenntnis (Klauer, 1977). Pädagogische Diagnostik umfasst mit ihren Instrumenten die betroffene Person selbst, das Umfeld der Person und insbesondere die schulische Situation. Die schulische Situation selbst kann von den Lehrkräften gestaltet werden. Hier passieren die entscheidenden Veränderungen. In der schulischen Situation wird die pädagogische Arbeit definiert. Daher sollten auch pädagogische Problemstellungen vorwiegend in der schulischen Situation selbst verbessert werden.

Eine missverstandene Diagnostik ist es, zu glauben, dass Diagnostik nur aus Gründen der Klassifikation durchführt wird. Diagnostik sollte nicht als reiner Verwaltungsakt des Schul- bzw. Bildungssystems gesehen werden, um eine Rechtfertigung für einen Nachteilsausgleich bzw. einen sonderpädagogischen Förderbedarf zu liefern. Hierbei würden die Zusammenarbeit, der Informationsaustausch und das präventive Handeln zugunsten von remedialer Förderung oder im schlimmsten Fall zugunsten der Zuschreibung von Störungsbildern ohne weitere Unterstützung vernachlässigt werden. Diagnostik ist ein Werkzeug, welches je nach Menschenbild, eigener Kompetenz und Handlungsmöglichkeiten verschieden durchgeführt und interpretiert wird. Um dieses Spannungsfeld zu veranschaulichen, wird eine Problemsituation in zwei Fällen dargestellt.
Fall 1 findet in einem präventiven Schulsystem statt, in welchem die Zusammenarbeit optimal abläuft:

> Der 7-jährige Schüler Angelo hat Schwierigkeiten im Fach Deutsch. Die Lehrerin Frau Müller sucht die Schulpsychologin Frau Hansen auf, nachdem Angelo trotz mehrerer Veränderungen ihres Unterrichts und ihrer Lernmaterialien weiterhin Schwierigkeiten im Bereich Lesen und Rechtschreiben hat. Die Schulpsychologin Frau Hansen berät Frau Müller und geht mit ihr alle Fördermaterialien und pädagogischen Maßnahmen durch. Nachdem Frau Müller bisher nur Arbeitsblätter und eigene Materialien zur Feststellung des Leistungsstandes eingesetzt hat, wird mit der ganzen Klasse ein Lesescreening durchgeführt, um in diesem Bereich die Förderung im Lesen für die gesamte Klasse zu verbessern. Nach Auswertung des Lesescreenings zeigt sich, dass Angelo eine geringe Geschwindigkeit in der Leseflüssigkeit hat. In diesem Bereich wird Angelo in den nächsten Wochen in Differenzierungsstunden, mit individualisierten Aufgaben im Wochenplan und bei den Hausaufgaben sowie im Klassenunterricht mit Lesepat:innen gefördert. Es besteht ein enger Kontakt zu den Eltern. In drei Monaten wird Angelo nochmals im Lesen mittels eines Screenings getestet. Frau Hansen bleibt in Kontakt mit Frau Müller und

dokumentiert den Lernprozess für sich. Frau Hansen unterstützt bei der weiteren Förderplanung und ist im Austausch mit allen Beteiligten. Regelmäßig findet im Schuleben ein Arbeitskreis zur Verbesserung der Förderung von Kindern mit Lernschwierigkeiten statt.

Ein solch präventives Vorgehen findet sich aktuell in der pädagogischen Praxis eher selten. Meist stehen einem präventiven Vorgehen Anforderungen und Regeln des Schulsystems, bürokratische und datenschutzrechtliche Bedenken sowie weitere Schwierigkeiten bei der Zusammenarbeit entgegen. Daher ist nachfolgend Fall 1 unter anderen Voraussetzungen als fiktiver **Fall 2** beschrieben, welcher in der Praxis wahrscheinlich vorkommt:

Der Schüler Angelo hat Schwierigkeiten im Fach Deutsch. Die Lehrerin Frau Müller interpretiert dies als eine Störung des Lesens, da er im Unterricht mehr Zeit als seine Mitschüler:innen benötigt, beim lauten Vorlesen stockend, leise und wenig betont liest und Aufgaben zum Textverständnis langsamer bearbeitet. Sie informiert die Eltern und holt deren Einverständnis ein, dass eine schulpsychologische Stellungnahme für einen Nachteilsausgleich verfasst wird. Die daraus resultierenden Folgen werden nicht erwähnt und auf das Gespräch mit der Schulpsychologin verwiesen. Die Mutter von Angelo denkt, dass Frau Müller die Diagnostik benötigt, um ihren Unterricht zu verändern, und das Konzept des Nachteilsausgleichs wurde ihr nicht erklärt.
Die Schulpsychologin Frau Hansen führt eine ausführliche Anamnese durch und testet Angelo an mehreren Vormittagen in den Bereichen der kognitiven Fähigkeiten, der Lesekompetenzen und der Wahrnehmung. Eine Aufklärung der Eltern über den Sachverhalt und die Folgen findet nur unzureichend statt. Angelos Mutter ist überrascht, als sie eine einseitige Stellungnahme mit der Feststellung einer Lesestörung bekommt. Als empfohlener Nachteilsausgleich darf die Arbeitszeit um ein Viertel verlängert werden. Ebenso sollen Aufgabenstellungen vorgelesen und der Lesetext vergrößert werden (individuelle Maßnahmen). Auf eine Benotung des Vorlesens kann verzichtet werden (Notenschutz). Empfehlungen für eine gezielte Förderung und weitere Maßnahmen der individuellen Unterstützung sind in der Stellungnahme nicht vermerkt. Ebenso findet kein fachlicher Austausch zwischen Schulpsychologin und Klassenlehrkraft statt.
Auf Nachfrage sendet die Schulpsychologin die Testergebnisse an Angelos Mutter, da diese darauf besteht, um für eine zweite Meinung einen Kinderarzt aufzusuchen. Ein Gespräch zwischen Eltern und Schulpsychologin findet aus zeitlichen Gründen nicht statt. Nach den Ergebnissen der Tests ist Angelo überdurchschnittlich intelligent, hat aber eine unterdurchschnittliche Leseflüssigkeit. Ein Gespräch zwischen Eltern und der Lehrkraft Frau Müller findet aus zeitlichen Gründen nicht statt. Frau Müller wechselt am Ende des Jahres in den Ruhestand und die Unterrichtszeit wird für das Üben der Abschlusstheateraufführung benötigt. Eine intensivere Förderung des Lesens findet nicht statt. Die Mutter ist verzweifelt. Sie hat der Diagnostik zugestimmt, in der Hoffnung, dass Angelo mehr oder bessere Unterstützung bekommt. Dem empfohlenen Nachteilsausgleich widerspricht sie. Von der Schule ist sie irritiert und insbesondere durch den Befund einer Störung sehr verunsichert. Über eine weitergehende Diagnostik zur Feststellung einer Lesestörung sowie einer Förderung durch Lerntherapie denkt sie nach.

Arbeitsauftrag:
Welcher pädagogischen Fragestellung geht die Schulpsychologin im Fall 1 und Fall 2 nach?
Welche Aufgaben und Rollen haben die Eltern im Fall 1 und Fall 2?
Welche Förderungen erhält Angelo im Fall 1 und Fall 2?

Weitere Informationen:
Handbuch des ISB Bayern (2019) zum Nachteilsausgleich

Arbeitsauftrag:
Was ist nach dem ISB Bayern (2019) individuelle Unterstützung? Nachteilsausgleich oder Notenschutz? Welches pädagogische Konzept oder Vorstellung von Lernen steht hinter der Definition nach dem ISB Bayern (2019)?

Während im Fall 1 eine Prozessdiagnostik mit enger Kooperation beschrieben ist, sind im Fall 2 die beteiligten Institutionen Schule und Schulpsychologie getrennt. Die Schulpsychologin führt im Fall 1 eine den Lernprozess begleitende Diagnostik durch und arbeitet eng mit der Klassenlehrkraft zusammen. Die Förderung erfolgt in der Schule. Die pädagogische Frage ist hierbei, wie man Angelo möglichst schnell und präventiv im Bereich Lesen helfen kann, bevor weitere schulische Probleme die Folge sein können. Im Fall 2 erfasst die Schulpsychologin im Rahmen einer Statusdiagnostik Angelos mögliches Anrecht auf einen Nachteilsausgleich. Die Frage der Förderung bleibt vom Schulsystem unbeantwortet und wird auf das Gesundheitssystem übertragen.

Das Fallbeispiel mit den zwei Wegen verdeutlicht, dass je nach Unterstützungssystem, Kooperation und Verständnis von schulischer Förderung eine andere Form, Fragestellung und Interpretation von pädagogischer Diagnostik praktiziert wird. Die Wirksamkeit der Diagnostik hängt eng mit der Umsetzung in Bereichen des schulischen Arbeitens zusammen und kann erst erfolgreich sein, wenn die im Rahmen der Diagnostik gewonnenen Erkenntnisse weitergegeben und angewendet werden. Arbeiten die beteiligten Personen nicht miteinander oder sogar über die Eltern und dem Kind hinweg statt gemeinsam, kann an dem Nutzen einer Diagnostik gezweifelt werden.

Take-Home-Message:
Pädagogische Diagnostik ist kein Selbstzweck. Sie ist mehr als ein Nachweis und eine Begründung einer Klassifikation. Erst im Zusammenspiel mit den pädagogischen Handlungen und deren Bewertung erhält die Diagnostik eine Berechtigung in pädagogischen Institutionen.

1.1 Diagnostik als Teil der inklusiven Schule

Welche Form der pädagogischen Diagnostik als Aufgabengebiet und Teil der Schule gesehen wird, ist eng verknüpft mit dem Rollenverständnis und den Aufgaben von Schule selbst. Was ist eine gute Schule? Wie inklusiv und adaptiv muss eine Schule sein? Dies sind politische und gesellschaftliche Fragen, welche sich im Wandel der Zeit immer wieder verändert haben.

Traditionell hat Diagnostik in der Schule eher einen geringeren Stellenwert (Klauer, 1977). Für den Unterricht genügte ein guter Standard für alle. Besonderheiten und Ausnahmen konnten und können nur gewährt werden, wenn nach dem individuell-medizinischen Modell eine Besonderheit wie beispielsweise eine kindorientierte (Lern- und Verhaltens-) Störung oder Beeinträchtigung festgestellt wird. In diesem Fall wird die Diagnostik benötigt, um eine Ausnahme von den bestehenden Regeln oder eine besondere Förderung zu ermöglichen.

Die Sichtweise, dass Schule und Unterricht für alle dieselben Inhalte zur gleichen Zeit mit denselben Methoden mit gleicher Prüfungsform vermitteln sollten, ist schockierender weise in der Praxis noch immer anzutreffen, aber nicht mehr Stand der Pädagogik und der empirischen Bildungsforschung. Für die inklusive Schule steht die Qualifizierung aller Schüler:innen im Vordergrund, mit dem Ziel, für alle Schüler:innen mit wirksamen Unterstützungsmethoden ein erfolgreiches akademisches Lernen und eine volle soziale Teilhabe (UN-BRK, 2006) zu ermöglichen. Dies bedeutet, dass die Verschiedenheit und die Leistungsheterogenität der Kinder als positive Bereicherung angesehen werden. Aufgrund von unterschiedlichem Vorwissen, Erfahrungen und individuellen Dispositionen lernen Kinder

unterschiedlich schnell und haben unterschiedliche Interessen und Stärken. In einer inklusiven Schule hilft die Diagnostik der Lehrkraft, die passende Lernumgebung zu gestalten und individuelles Lernen zu ermöglichen. Insbesondere mit Hilfe der Digitalisierung und computergestützten Test- und Lernumgebungen ist individuelles Lernen und Bewerten auch in Schulen immer leichter umzusetzen. Die Anwendung der schulischen Diagnostik unterscheidet sich daher, je nachdem ob ein selektives oder ein qualifizierendes Konzept verfolgt wird. Während selektive Konzepte ihren Schwerpunkt auf die Einteilung von Gruppen oder Klassifikationen legen, sind qualifizierende Konzepte auf die individuellen Entwicklung und dem Erreichen eines Bildungszieles fokussiert.

Eine Herausforderung des Bildungssystems ist, gleiche Chancen und Zugänglichkeiten für alle Personen zu ermöglichen. Spätestens seit dem PISA Schock ist auch in der Öffentlichkeit bekannt, dass dies dem deutschsprachigen Bildungssystem insbesondere bei Kindern aus bildungsbenachteiligten Familien nur unzureichend gelingt. Allen Personen ähnlichen Unterricht anzubieten und anhand von verschiedenen allgemeinen und besonderen Schularten zu unterscheiden hat keine gleichwertigen Chancen oder Zugänge von Bildung ermöglicht. Für ein inklusives Schulsystem bedeutet dies, dass alle Personen Zugang zu einer inklusiven Schule benötigen, in welcher je nach Bedürfnis möglichst präventiv unterstützt wird. Bei Kindern mit besonderen Bedürfnissen, bei auftretenden Lernschwierigkeiten, einer Behinderung oder anderen herausfordernden Gründen wird mehr Unterstützung und auch ein verzahntes Unterstützungssystem benötigt. Je nach Modell wird für 5% bis 25% der Schüler:innen besondere Unterstützung in einem mehrstufigen System benötigt. In einem solchen Modell wäre die Aufgabe der pädagogischen Diagnostik, die Verteilung der Unterstützung und die Wirksamkeit der Unterstützung durch die individuelle Lernentwicklung festzustellen (siehe für mehr Informationen hierzu Gebhardt, 2022). Es ist daher eine politische Frage, nicht nur welche und wie viele Ressourcen, sondern auch welche Kontrollen bereitgestellt werden, um auch Kindern mit Behinderung gleichwertige Bildungschancen zu ermöglichen, wie es die UN-Menschenrechtskonvention für Menschen mit Behinderung (UN-BRK, 2006) fordert.

In den USA wurde diese Frage vor Gericht im Fall Endrew F. v. Douglas County School District geklärt. Endrew (Drew) hat seit der Kindergartenzeit einen diagnostizierten sonderpädagogischen Förderbedarf (Special Educational Need) durch Autismus und einen sonderpädagogischen Förderplan (IEP):

„Drew attended Douglas County schools from preschool through fourth grade. During these years, the school district provided Drew with special education and related services as specified in his IEP.
Despite these services, Drew made minimum progress and continued having academic and behavioral challenges. As a result, Drew's parents placed him in a private facility specializing in educating students with autism, and rejected the district's IEP for fifth grade." (Prince et al., 2018, S. 322)

Die Eltern forderten die Kosten für die private Förderschule und weitere Förderungen vom Schuldistrikt. Dieser lehnte ab, da er eine staatliche sonderpädagogische Förderung ermöglichte. In den ersten Instanzen bekam der Schuldistrikt recht, jedoch nicht vor dem Supreme Court (Oberstes Gericht der USA). Hierbei war die entscheidende Frage, was eine „appropriate education" ausmacht und ob es ausreicht, dass sich die Förderziele von Endrew nicht wesentlich positiv oder nur minimal änderten. Für den Supreme Court war entscheidend, dass eine Förderung nicht trivial sondern bedeutungsvoll sein muss, um den Individuals with Disabilities Education Act nicht zu verletzen. Der Supreme Court legte fest, dass ein Förderplan nachvollziehbar berechnet und evaluiert werden muss, um für das Kind unter den individuell gegebenen Umständen Fortschritte zu ermöglichen (Prince et al., 2018).

Prince et al. (2018) empfiehlt daher den Praktiker:innen in den USA, dass Förderpläne (IEP)...

- den Bedarf mittels geprüfter Diagnostik feststellen,
- erreichbare und messbare jährliche Ziele haben,
- die gewählte sonderpädagogische Förderung sowie das Erreichen von Fortschritten umfassen und
- Instrumente zur Messung des Fortschritts wie Lernverlaufsdiagnostik (Progress Monitoring) beinhalten.

Weitere Informationen zum Fall:
https://www.law.cornell.edu/supct/cert/15-827

Kurzes Video zum Case Endrew

Ausführliches Video: Mitchell Yell, Ph.D. and David Bateman, Ph.D.

Arbeitsauftrag:
Holen Sie sich weitere Informationen zum Fall Endrew ein. Beschreiben Sie die Eckpunkte des Falles Endrew F. in eigenen Worten.
Welche Bedeutung hat die Entscheidung des obersten Gerichts für die sonderpädagogische Förderung in den USA? Welche Veränderung bedeutet dies für die pädagogische Praxis?
Warum ist nach Prince et al. (2018) die Evaluation und die Messung des Fortschritts notwendig?

In Deutschland hat das Urteil des Obersten amerikanischen Gerichts keine unmittelbaren Folgen. Trotzdem stellt sich die Frage, wie ein solcher Fall vor deutschen Gerichten ausgehen würde. Ein minimaler Standard genügt eher nicht, um die Ziele der Inklusion umzusetzen, so wie es die UN-Behindertenrechtskonvention (2006) fordert. Wenn die Rechte aller Schüler:innen ernst genommen werden, muss es um eine nachweisliche „appropriate education" gehen. Bis zu dem Urteil war es in den USA ähnlich wie in Deutschland möglich, nur auf den angebotenen Unterricht und die bestehenden Fördermaßnahmen zu verweisen, ohne die Unterstützungsmaßnahmen explizit anhand der konkreten individuellen Bedürfnisse der lernenden Person zu evaluieren und den Entwicklungsverlauf zu begründen.

Take-Home-Message:
Während es in den USA ein einklagbares Recht auf die Durchführung und Evaluation von appropriate education für Menschen mit Behinderung gibt, ist dies in Deutschland aktuell nicht der Fall.

Arbeitsauftrag:
Wie könnte so ein einklagbares Recht auf die Durchführung und Evaluation von appropriate education für Menschen mit Behinderung in Deutschland aussehen und welche Folgen und Anpassung würde dies für das Schulsystem und einzelne Schulen bedeuten.

In Gerichtsurteilen wie im Fall Endrew F. vs. Douglas County School District werden die diagnostische Praxis und die Interventionen anhand der Dokumentation und der Gutachten geprüft. Hierfür ist es nötig, dass es Standards und Klassifikationen gibt, welche bei der Dokumentation eingehalten werden. Die Dokumentation besteht dabei aus den Ergebnissen der Beobachtungen und der eingesetzten Instrumente, welche in Form von qualitativen (Wortgutachten) und quantitativen Daten (Tests und Beobachtungen) interpretiert werden. Hierbei müssen neben individuellen Beobachtungen auch standardisierte Instrumente eingesetzt werden, welche eine nachweisliche Güte haben. Nur so kann später nachvollzogen werden, dass die getroffenen Entscheidungen nicht willkürlich waren und eine Vergleichbarkeit gegeben ist. Eine vollständig individuelle Diagnostik ohne eine Form von

Standardisierung und ohne ein geprüftes Instrument ist für die pädagogische Praxis nicht zielführend. Eine solche Diagnostik wäre sehr aufwendig, im hohen Maße von den Kompetenzen und Vorstellungen der diagnostizierenden Personen mitbestimmt und in ihren Ergebnissen kaum vergleichbar. Pädagogische Diagnostik beinhaltet daher sowohl individuelle Beschreibungen als auch mindestens eine vergleichbare Messung und setzt sich aus qualitativen und quantitativen Daten zusammen.

Vor dem Gerichtsurteil von Fall Endrew F. stand bei der Diagnostik fast ausschließlich der aktuelle Lern- und Verhaltensstatus im Vergleich zu allen Kindern im Vordergrund, um daraufhin in Form eines Förderplans Interventionen abzuleiten. Dies wird als **Statusdiagnostik** bezeichnet. Die alleinige Messung eines Lernstandes genügt aber nicht, um Lernentwicklung und schulischen Erfolg nachzuweisen. Daher wird nun die **Prozessdiagnostik** beispielsweise in Form der **Verlaufsdiagnostik** wichtiger, um den Fortschritt und den Erfolg des Lernens auf individueller Ebene am Kind mittels mehrerer Messungen über einen größeren zeitlichen Abstand zu dokumentieren. Dies wird somit auf lange Sicht die schulische Diagnostik verändern. Diagnostik wird in der Schule kein einmaliges Ereignis sein, welches als externe Dienstleistung erbracht wird, sondern eine eigenständige Aufgabe aller Lehrkräfte, welche begleitend zum Unterricht stattfindet. Diagnostische Prozesse sowie Unterrichtsplanung, -durchführung und -evaluation hängen somit enger zusammen. Diagnostik wird für alle Schüler:innen immer dann eine Rolle spielen, wenn es um das Erreichen wichtiger Kompetenzen und langfristiger Ziele geht.

Im Schulsystem werden Schulen als Verwaltungseinheiten betrachtet. Für einzelne Personen kann ein Antrag über eine besondere Leistung (auf eine sonderpädagogische Unterstützung) gestellt und gewährt werden. Aktuell genügt es dem Schulsystem, wenn eine Maßnahme erteilt und deren Inhalt vorher definiert wurde. Dies wird als Input Steuerung bezeichnet. Wird neben dem Erhalt der Förderung auch die Frage der Wirksamkeit auf der Ebene des Kindes betrachtet, wird der Output des Systems bewertet. Diese Frage nach der Wirksamkeit von Pädagogik im Einzelfall ist für das Schulsystem immer noch ein neues Feld. Während für die empirische Forschung Fragen zur Wirksamkeit und Evidenz insbesondere seit der empirischen Wende wesentlich sind (siehe Kapitel zur Evidenz im Einführungsbuch), spielen solche Fragen im Schulsystem nur eine untergeordnete Rolle, selbst bei gravierenden Entscheidungen wie beispielsweise Einschulung, Klassenwiederholung, Schulwechsel oder Zuweisung eines sonderpädagogischen Förderbedarfs oder dem Verfassen eines rechtssicheren Gutachtens. Die Wirksamkeit von Schule und schulischer Förderung auf der Ebene des Kindes wird aktuell in Deutschland weder gesetzlich gefordert noch im System nachgeprüft. Je nach politischen Vorgaben wird bei der Diagnostik in der Schule in Zukunft entweder die Statusdiagnostik zur Erfassung eines aktuellen Ist-Standes oder die Prozessdiagnostik, zur Betrachtung einer Lern- oder Verhaltensentwicklung im Vordergrund stehen. Bislang ist ersteres der Fall.

Take-Home-Message:
Für die sonderpädagogische Förderung wird die Wirksamkeit von pädagogischen Maßnahmen für den Einzelfall zunehmend von Bedeutung sein.

Diagnostik beantwortet systemisch und analytisch eine pädagogische Fragestellung. Die Breite diagnostischer Fragestellungen kann von einzelnen Fragen zu bestimmten didaktischen Entscheidungen oder bis zu gravierenden Entscheidungen zum Wechsel einer Schule oder dem Besuch von speziellen Fördereinrichtungen reichen. Je nach Konsequenz, Umfang der Förderung und Kosten für das Schulsystem sollten Umfang und Aufwand verhältnismäßig sein und benutzerfreundliche, aber reliable, faire und zuverlässige Instrumente verwendet werden. Aktuell ist das Kosten-Nutzen-Verhältnis in unserem Schulsystem meist nicht gegeben, da eine aufwändige Statusdiagnostiken benötigt wird, um wenige remediale Förderstunden zu erhalten. Hoch wirksame präventive Förderung anhand von punktuellen, aber reliablen Instrumenten kommt dagegen meist zu kurz oder ist in der

Praxis unbekannt. Daher ist es eine offene Frage, ob eher eine präventive oder eine remediale Praxis mit schulischer Diagnostik verfolgt wird.

1.2 Pädagogische Diagnostik

Lehrkräfte erfassen und bewerten permanent in ihrem Unterrichtsalltag und nehmen damit diagnostische Handlungen vor, welche Sie für ihre Handlungsentscheidungen nutzen. Es kann somit nicht nicht diagnostizieren werden, da Annahmen über den Leistungsstand, das Vorwissen, die kognitiven Fähigkeiten und die mögliche Lernentwicklung getroffen werden, wenn Unterrichtsreihen und Lernumgebungen geplant werden. Pädagogische Diagnostik ist Aufgabe aller Lehrkräfte und Teil ihrer Arbeit. Die Standards der Lehrer:innenbildung nach der KMK (2004) formulieren diese Kompetenz als die Aufgabe, Lernvoraussetzungen und Lernprozesse von Schüler:innen zu diagnostizieren, um die Schüler:innen gezielt zu fördern und zu beraten. Diagnostik ist daher Teil der Schule und kann je nach Lehrkraft sehr unterschiedlich aussehen. Neben der förderorientierten Diagnostik gibt es auch noch umfangreichere Diagnostik im Rahmen von schulpsychologischen oder sonderpädagogischen Fragestellungen. Die Methodik und die Grundsätze von Diagnostik bleiben aber gleich (Gebhardt, Scheer & Schurig, 2022).

Missverständnisse von pädagogischer Diagnostik:
Ein im schulischen Alltag anzutreffendes Missverständnis ist, dass Diagnostik nicht als Erkenntnisprozess für pädagogische Entscheidungen im pädagogischen Alltag erkannt wird, sondern nur als von externen Einrichtungen oder der Schulpsychologie angebotene umfangreiche Einzeldiagnostik wahrgenommen wird. Pädagogisch diagnostisches Arbeiten ist jedoch mit jedem Erkenntnisgewinn über den aktuellen Lernstand, die Lernentwicklung und die bestmögliche individuelle Förderung für die Vorbereitung und Durchführung einer Lern- und Unterrichtsumgebung verbunden. Diagnostisches Arbeiten findet somit implizit oder explizit immer statt, die Frage ist eher mit welchen Instrumenten und wie umfänglich.

Auch in der Lehre sind die ersten Fragen von Studierenden meist auf das Erkennen von Störungsbildern gerichtet. Erst im weiteren Verlauf des Studiums wird die Verknüpfung von Diagnostik, Handlungsentscheidung und eigenem pädagogischem Handeln reflektiert.

Arbeitsauftrag:
Befragen Sie eine Lehrkraft, die Sie kennen, was sie genau unter Diagnostik versteht und welche diagnostischen Methoden Teil ihrer Arbeit sind.

Pädagogische Diagnostik unterscheidet sich auch im Ablauf und der Methodik nur gering von psychologischer Diagnostik. Der Unterschied liegt in der Art der Fragestellung und darin, welchen Hintergrund die durchführende Person hat. So findet eine klinische psychologische Diagnostik meist in Einrichtungen des Gesundheitswesens statt und bewertet eine klinische Fragestellung beispielsweise nach Kriterien der ICD 11. Psychologische Diagnostik beantwortet Fragen zu den Bereichen der Beschreibung, Klassifizierung, Erklärung und Vorhersage menschlichen Verhaltens und Erlebens (Schwaighofer et al., 2019). Unter Klassifizierung wird dabei verstanden, die Person in eine Klassifikation oder ein Raster wie beispielsweise die ICD 11 einzuordnen. Die Klassifizierung ist jedoch nur bei bestimmten Fragestellungen ein notwendiger Teil des diagnostischen Prozesses. Das eigentliche Ziel des diagnostischen Prozesses ist nicht die Einordnung in ein Raster, sondern der Vorschlag der helfenden Intervention (Kubinger, 2009).

Die pädagogische Diagnostik dagegen bezieht sich auf den Kontext der Pädagogik und klärt diagnostische Fragen rund um diesen Bereich. Klauer sieht pädagogische Diagnostik daher sehr allgemein und je nach Fragestellung variabel:

"Pädagogische Diagnostik ist das Insgesamt der Erkenntnisbemühungen im Dienste aktueller pädagogischer Entscheidungen" (Klauer, 1977, S. 5).

Für Klauer steht damit die Bemühung des Erkennens im Vordergrund. Im Unterschied zur Wissenschaft, welche sich um allgemeine Erkenntnisse bemüht, liegt der Fokus in der pädagogischen Diagnostik auf dem **Einzelfall**. Pädagogische Diagnostik findet immer im Sinne von Planungsentscheidungen, Handlungsentscheidungen oder Feststellungs- und Bewertungsentscheidung (Beurteilung) statt (Klauer, 1977).

Diagnostik ist daher ein Bemühen um Erkenntnis, um die bestmögliche Entscheidung zu treffen. Durch die methodische Beschreibung der Problemstellung und der Beantwortung einer Frage soll auf den Kern des Problems geschlossen und Handlungsempfehlungen gegeben werden. Diagnostik ist daher kein Selbstzweck, sondern Teil der Lösung einer Aufgaben- und Problemstellung.
In der Praxis ist es meist systemisch oder institutionell bedingt, dass diagnostische Methoden nur noch als Abfolge benutzt werden, um ein bestimmtes Vorgehen zu rechtfertigen oder ein System zu stützen. Dies wird dann zwar dem Namen nach Diagnostik genannt, hat aber mit der eigentlichen Idee, die hinter Diagnostik steckt, nichts zu tun. Denn wenn die Diagnostik Selbstzweck des Systems und nicht für den betroffenen **Einzelfall eine Hilfe** ist, ist sie keine pädagogische Diagnostik mehr. Aus diesem Grund muss die Diagnostik zuerst vom individuellen Kind und im zweiten Schritt vom System Schule ausgehen. Eine Voraussetzung für gute Diagnostik ist, den Blickwinkel des Kindes und seines Umfelds einzunehmen, Probleme zu identifizieren und dann Lösungen für das System Schule zu finden.

Take-Home-Message:
Diagnostik ist ein Bemühen um Erkenntnis, um die bestmögliche Entscheidung zu treffen. Je nach Ansatz und Fragestellung ist die jeweilige Diagnostik sehr unterschiedlich.

Im Gegensatz zu Klauer definieren Ingenkamp und Lissmann (2008) die pädagogische Diagnostik in Bezug auf bestimmte pädagogische Aufgaben pragmatischer. Hierbei haben Ingenkamp und Lissmann (2008) die Anforderungen der allgemeinbildenden Schule im Blick:

„Pädagogische Diagnostik umfasst alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren. Zur Pädagogischen Diagnostik gehören ferner die diagnostischen Tätigkeiten, die die Zuweisung zu Lerngruppen oder zu individuellen Förderungsprogrammen ermöglichen sowie die mehr gesellschaftlich verankerten Aufgaben der Steuerung des Bildungsnachwuchses oder der Erteilung von Qualifikationen zum Ziel haben." (Ingenkamp & Lissmann, 2008, S. 13)

Ingenkamp und Lissmann (2008) decken mit ihrer Definition einen großen Teil der Aufgaben der pädagogischen Diagnostik ab, welche in der allgemeinbildenden Schule vorkommen. Jedoch sind die Aufgabengebiete der pädagogischen Diagnostik in einer inklusiven Schule weitreichender und umfassen beispielsweise auch formatives Assessment und Evaluation von Maßnahmen sowie sonderpädagogische und schulpsychologische Fragestellungen. Ebenso kann Diagnostik als Teil der Schulentwicklung und Schulkultur betrachten werden. Darüber hinaus gehört zu den weiteren Feldern der pädagogischen Diagnostik das Erstellen und die Evaluation von Instrumenten und Tests sowie die Planung, Erstellung, Durchführung und Evaluation von Interventionen und Maßnahmen. Daher ändern sich die Aufgabengebiete der pädagogischen Diagnostik, je nachdem wie Schule gesehen und definiert wird. Die allgemeine Definition von Klauer (1977) trifft daher die diagnostischen Schritte besser.

Arbeitsauftrag:
Erörtern Sie die Unterschiede und Gemeinsamkeiten der Definitionen von Klauer (1977) und Ingenkamp und Lissmann (2008).

Als unterschiedliche Strategien der Diagnostik kann die Selektion, die Modifikation oder auch die Veränderungsmessung gesehen werden. Pawlik (1982) unterscheidet hier zwischen der Selektion und der Modifikation. **Selektion** ist eine einmalige Entscheidung, welche zu einem Zeitpunkt erfolgt, um sich zwischen zwei oder mehreren Alternativen zu entscheiden. Dies ist beispielsweise die Entscheidung zwischen Förderschule und Grundschule oder auch den verschiedenen Schularten der Sekundarstufe. Es handelt sich um eine Entscheidung zu einem bestimmten Zeitpunkt. Im Kern steht die Frage, zu welchen Anforderungen der Schule das Kind passt. Hierfür wird eine umfangreiche Statusdiagnostik zu einem Zeitpunkt benötigt, um die zukünftige Entwicklung des Kindes mit den Anforderungen der Schule zu vergleichen und den „passenden Schulort" zu wählen.

Die **Modifikationsdiagnostik** stellt die Frage nach der Auswahl der Intervention oder Lehrmethoden. Die Modifikationsdiagnostik bezieht sich nach Pawlik (1982) auf die Verhaltenstherapie, bei welcher die Veränderung des Verhaltens im Vordergrund steht. Übertragen auf die Schule wird bei der Modifikationsdiagnostik die Frage gestellt, wie der aktuelle Unterricht angepasst werden muss, um das Lernziel zu erreichen bzw. welche Förderungen notwendig sind.

Beispiel Schuleingangsuntersuchung:
Die Schuleingangsuntersuchung, durchgeführt von den Landesämtern für Gesundheit, ist eine flächendeckende differenzielle Untersuchung. Anhand von standardisierten Tests führt eine Fachkraft der Sozialmedizin (meist eine Kinderkrankenschwester) eine Untersuchung durch und gibt Eltern eine Einschätzung anhand eines Entwicklungsstandes als zusätzliche Entscheidungshilfe zur Einschulung. Die Eltern werden aus Sicht der Schulmedizin beraten und an weitere mögliche Beratungsstellen verwiesen. Die Daten werden nicht an Schulen weitergegeben. Daher findet bei der Einschulung meist eine eigene Form der Einschulungsdiagnostik beispielsweise als sogenanntes Schulspiel statt.

Arbeitsauftrag:
Welche Form der Diagnostik beinhaltet die Schuleingangsuntersuchung? Beschreiben Sie diese in wenigen Sätzen.

Im Bereich der Schule gibt es als Form der Modifikation zwei unterschiedliche Konzepte: den älteren und verbreiteten Aptitude-Treatment-Interaktion-Ansatz (Cronbach & Snow, 1977) sowie den neueren Response-to-Intervention-Ansatz (Deno, 1990). Beide Konzepte unterscheiden sich maßgeblich durch den Einsatz der Diagnostik. Im Aptitude-Treatment-Interaktion-Ansatz werden vor und nach der Förderung summative Bewertungen zu maximal zwei Messzeitpunkten und beim Response-to-Intervention-Ansatz formativ begleitende Bewertungen zu mehreren Messzeitpunkten während des Lernens genutzt.

Der **Aptitude-Treatment-Interaction-Ansatz (ATI)** beschreibt das Lernen als Wechselwirkung zwischen den individuellen Merkmalen der Lernenden und der unterrichtlichen Förderung (Tröster, 2019). Zuerst konzentrieren sich die Lehrkräfte auf die Merkmale der Lernenden (Aptitude) und wählen anschließend eine passende Intervention (Treatment) aus und bieten sie den Lernenden an. Grundannahme ist, dass jede:r Lernende unterschiedliche Hilfen benötigt. Eine Umsetzung des ATI-Ansatzes ist ein adaptiver Unterricht, in dem zuerst eine summative Leistungsfeststellung bei allen Schüler:innen zu einem Zeitpunkt erfolgt. Im Anschluss werden die Schüler:innen in verschiedene Gruppen aufgeteilt und nach diesen Gruppen geordnet gefördert. Nach dem Lernprozess werden dann die Kompetenzen der Lernenden ein zweites Mal summativ evaluiert. Diese Überprüfungsart wird summative Evaluation genannt und findet jeweils vor und nach der Förderphase mit einem Test statt. In einer summativen Evaluation ist erst nach der Förderung zu erkennen,

ob das Ziel erreicht wurde. Dieses Vorgehen entspricht dem forschungsbasierten Vorgehen in der pädagogisch-psychologischen Forschung, um die Wirksamkeit von Interventionen in einem experimentellen Forschungsdesign festzustellen (Blumenthal et al., 2014). Auf diesen Ansätzen bauen die Konzepte im adaptiven Unterricht auf. Für dieses Vorgehen in der Diagnostik gibt es meist eine Statusdiagnostik für alle Personen, auf deren Grundlage dann die Einteilung zu bestimmten passenden Interventionen erfolgt. Nach der Intervention erfolgt eine weitere Testung, um die Fortschritte zu evaluieren.

Walter (2008) und Blumenthal et al. (2014) kritisieren am ATI-Ansatz, dass dieses Vorgehen mit einer Einteilung nach den Ist-Leistungsständen von Schüler:innen auf Grundlage von einmaligen summativen Leistungsmessungen oder informellen Beobachtungen für individuelle Förderentscheidungen nicht ausreicht. Diese Art der Evaluation genügt nicht den Ansprüchen eines individualisierten Unterrichts in heterogenen Lerngruppen. Walter (2008) kritisiert außerdem, dass sowohl Forschende als auch Lehrkräfte ihr Denken und Handeln meist unreflektiert nach diesem Paradigma gestalten. Als Alternative sehen sie den präventiv orientierten US-amerikanischen Response-to-Intervention-Ansatz.

Im Fokus des **Response-to-Intervention-Ansatzes (RTI)** stehen die Lernentwicklungen einzelner Schüler:innen. Es wird der Frage nachgegangen, ob die Förderung auf das Lernziel wirkt oder nicht (Fuchs & Fuchs, 2006). Zur Beantwortung dieser Frage werden die Lernentwicklungen der Schüler:innen mit der Lernverlaufsdiagnostik mittels formativer Beobachtungs- und Testdaten im Längsschnitt begleitet und evaluiert. Für die Evaluierung der aktuellen Förderentscheidungen werden also mehrere Messungen und Informationen während des Lernprozesses erhoben, da zwei Prä-Post-Messungen mit normierten Schulleistungstests für didaktische Entscheidungen eine unzureichende Datengrundlage darstellen (Fuchs et al., 1983). Anschließend werden auf Grundlage der Daten Entscheidungen über mögliche Anpassungen in der Förderung getroffen. Vaughn et al. (2003) sehen gerade in der Verwendung der Daten zur Lernentwicklung den generellen Paradigmenwechsel weg von der Feststellungsdiagnostik hin zur Förderdiagnostik. Im ATI-Ansatz wird ein großer Teil der Zeit in die summative Erhebung des Ist-Standes und die Zuteilung nach Gruppen investiert. Der RTI-Ansatz ist dagegen begleitend und kann durch die formative Erhebung einzelne Fördermaßnahmen pro Person anpassen und verändern. Beide Ansätze wählen verschiedene Förderungen aus. Der Unterschied liegt hauptsächlich darin, ob die Erhebung mittels summativer oder formativer Erhebung erfolgt. Mit dem RTI-Konzept ist neben der veränderten diagnostischen Grundhaltung ein Mehrebenenpräventionsmodell, das Multi-Tiered Systems of Support (MTSS), verbunden, welches alle in Deutschland eher getrennten Systeme (Förderschule, Grundschule & Schulpsychologie) in ein Verteilungsmodell von Lehrer:innenstunden und Finanzen bündelt.

Take-Home-Message:
Während summative Evaluation erst am Ende der Maßnahme erfolgt und Informationen zur Wirksamkeit bietet, findet formative Evaluation begleitend zur Maßnahme statt und ermöglicht Informationen zu möglicherweise notwendigen Anpassung in der Maßnahme.

Pädagogische Diagnostik betrifft alle pädagogischen Fragestellungen und Entscheidungen im Alltag. Entscheidungen können daher für das Lernen und für Förderungen getroffen werden. Dieser Ansatz wird in der Schulforschung als **classroom-based assessment (CBA)** bezeichnet, in welchem es um pädagogische und didaktische Entscheidungen für den Unterricht geht (Genesee & Hamayan, 1994). Auch hier geht es um Verfahren der Status- und Prozessdiagnostik, wobei die Prozessdiagnostik im Vordergrund steht. Davon abzugrenzen ist eine umfangreichere differentielle Diagnostik zur Klärung im Bereich der Schulpsychologie oder Sonderpädagogik, in welcher es um weitreichende Beratungen zur Schulkarriere oder Schulorten geht. Jungjohann und Gebhardt (im Druck) sehen daher beispielsweise für Sonderpädagog:innen vier Dimensionen für die diagnostische Kompetenz als notwendig an, welche mittels des Fragebogens DAKI geprüft werden können:

- Fachdidaktische Diagnosekompetenzen zur Planung von gutem Unterricht
- Kompetenzen im Bereich Educational Assessment zur Durchführung und Interpretation von Diagnostik zum Status Quo
- Kompetenzen im Bereich Gutachtenerstellung und Feststellung von sonderpädagogischen Unterstützungsbedarfen
- Kompetenzen im Bereich der Feststellung von Lernfortschritten und Lernverlaufsdiagnostik

Videos:
Diagnostik in der Schule (25:15)
Bezugsnormenorientierung (14:17)

Arbeitsauftrag:
Beschreiben Sie die Diagnostik in eigenen Worten. Wann und wo wird Diagnostik in der Schule verwendet? Was ist das Ziel von Diagnostik? Was wäre ein nicht konformes Vorgehen mit pädagogischer Diagnostik? Was ist der ATI-Ansatz? Was ist der RTI-Ansatz? Was sind die jeweiligen Unterschiede des ATI- und RTI-Ansatzes?

Weitere Informationen finden Sie im Handbuch zur sonderpädagogischen Diagnostik:
Gebhardt, M., Scheer, D. & Schurig, M. (2022). Sonderpädagogische Diagnostik. Eine Einführung. In. M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 7-15). Universität Regensburg.

Take-Home-Message:
Je nach pädagogischer Institution und ihrem jeweiligen Auftrag wird diagnostisches Handeln sehr unterschiedlich definiert. Dadurch werden je nach Institution und Zielgruppe unterschiedliche Instrumente und Herangehensweisen benutzt. Gemeinsames Ziel aller diagnostischen Prozesse ist der pädagogische Nutzen der daraus folgenden Handlung und die Begründung der pädagogischen Handlungsweisen und -entscheidungen vor sich selbst und gegenüber Außenstehenden. Daher ist der Aufbau eines Gutachtens und Förderplans häufig ähnlich strukturiert und standardisiert.

1.3 Didaktik und Diagnostik

Die Schule hat die Aufgabe alle Kinder zu bilden und eine optimale Entwicklung zu ermöglichen. Dies gelingt nach den verschiedenen OECD Studien PIRLS oder PISA in Deutschland im Vergleich zu anderen Ländern nur bedingt. Der soziale Gradient also der Zusammenhang zwischen dem sozialen Status und dem Kompetenzniveau ist in Deutschland höher als in anderen Staaten.

Ein Elternhaus mit hohem Bildungsgrad, ökonomischen Ressourcen hat mehr Möglichkeiten in außerschulische Bildungsmöglichkeiten zu investieren, kann das Kind optimaler auf den Unterricht und die schulischen Erwartungen vorbereiten und bei Konfliktfällen besser ihr eigenes Recht argumentieren und durchsetzen als andere Personengruppen (bsp. mit Fluchthintergrund, Armut, Behinderung und anderer Benachteiligung). Ein gut gestaltetes Bildungssystem kann diesen Vorteil des Elternhauses mit hohem Bildungsgrad zwar nicht vollständig ausgleichen, aber durch faire Bedingungen und Unterstützungssysteme abfedern. Inwieweit ein inklusiveres oder im Gegensatz dazu ein segregatives Bildungssystem gewünscht ist, ist eine gesellschaftliche und auch politische Frage. Mit dieser Ausrichtung des Bildungssystems hängt auch der Einsatz und die Art der pädagogische zusammen. So kann Diagnostik einsetzt werden, um Sondergruppen zu definieren und Unterschiede zu beleuchten oder um Barrieren im Bildungssystem abzubauen und Zugänge zu ermöglichen.

Pädagogische Diagnostik ist daher ein Baustein im Bildungssystem, der um so wirkungsvoller ist, umso mehr er mit den pädagogischen und didaktischen Entscheidungen verknüpft ist. Dieser Zusammenhang ist beidseitig, denn eine diagnostische Erkenntnis kann umso genauer und detaillierter sein, umso mehr über die verschiedenen pädagogischen Maßnahmen der durchführenden Diagnostiker:in bekannt ist. Auf der anderen Seite verbessert sich eine Förderung, umso mehr durch die diagnostische Erkenntnis über die Bedürfnisse der Person bekannt ist, denn dadurch kann möglichst individuell die Förderung an deren Bedürfnisse angepasst wird. Pädagogische Diagnostik ist daher ein Bestandteil von Schule und Unterricht, welcher einerseits flächendeckend und niederschwellig und andererseits individuell und spezifisch ist, um Prävention, Intervention und Rehabilitation zu ermöglichen.

Take-Home-Message:
Diagnostik und Assessment stellt eine Passung zwischen Fähigkeiten der Person und dem Anspruch der Förderung her. Ziel ist die optimale Lernumgebung, welche zu einer individuell optimalen akademischen und sozialen Entwicklung bei vollständiger Teilhabe ermöglicht.

1.4 Der diagnostische Prozess

Der diagnostische Prozess wird von Jäger und Petermann (1999) mit den drei Stufen Sammeln von Informationen, Verarbeiten von Informationen und Interpretation von diagnostisch relevanten Informationen beschrieben. Informationen werden je nach Einzelfall verschieden gesammelt und bewertet. Im Unterschied zu einer sogenannten Bauchentscheidung muss, ähnlich wie beim wissenschaftlichen Arbeiten, die Entscheidung begründet werden und nachvollziehbar sein. Daher werden drei Schritte im Ablaufprozess der pädagogischen Diagnostik Beschreiben und Beobachten, Bewerten der aktuellen Lage und Erklären, Interpretieren und Vorhersagen (Jäger & Petermann, 1999).

Diese drei Schritte durchläuft jede pädagogische Entscheidung implizit oder explizit. Im Alltag vertrauen viele Lehrkräfte auf ihr Bauchgefühl, daher wird nach einer kurzen Beobachtung direkt bewertet und interpretiert. Wird die Ausgangslage jedoch nicht systematisch, vergleichbar und nachvollziehbar erfasst, besteht die Gefahr von Beobachtungs- bzw. Beurteilungsfehlern. Daher ist die Beurteilungskompetenz und damit verbundene Urteilsgenauigkeit von Lehrkräften meist durchaus valide. Sie können die Schüler:innen meist treffsicher ordinal in eine leistungsstarke, mittelstarke oder eher leistungsschwache Gruppe einteilen. Eher größere Schwierigkeiten haben Lehrkräfte beim Erkennen eines individuellen Profils zur Ableitung von individuellen Förderzielen, da die Lehrkräfte meist nicht systematische die Lernausgangslage erfassen (Kilday et al., 2012). Wird dies auf Personen mit einem spezifischen Merkmal wie einer Beeinträchtigung bezogen, welche außerhalb einer von der Schule gesetzten Norm oder eines Bewertungsrasters sind, können die Bedürfnissen dieser Personen von den Lehrkräften in ihrem Alltag nicht korrekt erkannt werden. Daher benötigt das Schulsystem diagnostische Prozesse im Unterricht, wie es beispielsweise im Team der Lehrkräfte einer Klassenstufe organisieren werden könnte, um Kinder mit Lernschwierigkeiten in bestimmten Fächern zu identifizieren und spezifisch zu fördern. Auf zweiter Ebene werden Expert:innen im diagnostischen Prozess gebraucht, um einerseits spezifische Wissen und Kompetenzen einzubringen und um andererseits eine externe und neutralere Einschätzung zu bekommen.

Exkurs:
Auch bei Erzieher:innen ist die Urteilsgenauigkeit ähnlich ordinal wie bei den Lehrkräften, wie die Studie von Donie (2021) zeigt. Die in Open Access erschienene Doktorarbeit stellt auch den aktuellen Forschungsstand zur Urteilsgenauigkeit sehr gut da.

Die Schritte im Prozess der Diagnostik helfen den diagnostischen Prozess zu strukturieren und vorschnelle Bewertung oder Interpretationen zu vermeiden.

Im **ersten Schritt Beschreiben und Beobachten** wird die für die aktuelle Problemstellung relevante aktuelle Ausgangslage deskriptiv und ohne Wertung zu erfasst, um ein umfassendes und wesentliches Bild der Problemstellung zu haben. Zu diesem Schritt gehören Anamnese, Gespräche mit allen Beteiligten, Beobachtungen, aber auch die Erfassung von schriftlichen Leistungen, Hausaufgaben und weiteren relevanten Quellen. Nach der Erfassung der Problemlage wird ähnlich wie im wissenschaftlichen Arbeiten eine **Fragestellung** und mögliche Hypothesen aufgestellt. Zur Beschreibung der Person gehören auch die Auswahl und Durchführung für die Problemstellung relevanter Tests und Fragebögen, um objektiv vergleichbare Daten auf Ebene der Person zu bekommen. Da Testdurchführungen organisiert werden müssen und die Kinder die Testsituation als belastend empfinden können, werden möglichst wenig aber doch ausreichend viele Tests ausgewählt (Siehe zur Testauswahl 6.1 und 6.3).

Handelt es sich beispielsweise um eine schulische Schwierigkeit in Mathematik liegen die Informationen in Hausaufgaben, Übungsblättern, Proben oder als ordinal gestufte Note vor. Diese Informationen sind jedoch nicht unbedingt systematisch aufbereitet und decken auch nur einen Teil der mathematischen Basiskompetenzen bzw. der Vorläuferfähigkeiten ab. Möchte man aber eine Übersicht über die mathematische Kompetenz haben, werden alle Aufgaben in einer systematischen Aufstellung benötigt, um eine solchen Kompetenz zu messen sowie einen Vergleichsmaßstab einer Alterskohorte benötigt. Diese Informationen bieten nur Tests (siehe Kapitel 5) als Gruppen- oder Einzeltest ökonomisch. Liegen alle relevanten Informationen vor ist der Schritt der Beschreibung und Beobachtung vorbei. In diesem Schritt erfolgt keine Bewertung.

Im **zweiten Schritt Bewerten der aktuellen Lage** wird geprüft, ob alle wesentlichen Informationen zur Beantwortung der Fragestellung vorliegen. Die einzelnen Informationen werden systematisch dargestellt, miteinander verknüpft und nach ihrer Aussagekräftigkeit gewichtet. Stimmen die Aussagen der betroffenen Personen überein? Zeigen die Ergebnisse ein ähnliches Bild oder widersprechen sich die Ergebnisse? Wie sind die eigenen Beobachtungen zu bewerten? Welche der aufgestellten Hypothesen kann verworfen werden?

Im dritten Schritt **Erklären, Interpretieren und Vorhersagen** wird anhand der Bewertung die Fragestellung beantwortet. Hierzu werden alle Informationen und Bewertungen herangezogen und aus dem Blickwinkel der Person als auch der Schule interpretiert. Die Ursachen des Problems werden erklärt und hierbei auch angegeben, für wie wahrscheinlich diese Interpretation gehalten wird. Hierbei wird vorherige Beschreibung des Status (Ist-Stand), des individuellen Profils mit Stärken, Schwierigkeiten, Interessen sowie der besonderen Bedürfnisse der Person in Bezug auf die Möglichkeiten und Anpassungsfähigkeit der der Bildungsinstitution gegenübergestellt und interpretiert. Für die pädagogische Diagnostik ist es nun wesentlich, dass aus diesem Status der Person und den Gegebenheiten in der Bildungsinstitution Vorhersagen und Empfehlungen gegeben werden. Unter einer Vorhersage versteht man, dass unter den gegeben und unter geänderten Umständen die weitere Entwicklung prognostiziert wird. Die Empfehlungen sollten so detailliert und konkret wie möglich beschrieben und kommuniziert werden, damit diese auch im pädagogischen Alltag umgesetzt werden können. In Fragenstellungen mit der Festsetzung eines Nachteilsausgleichs oder eines sonderpädagogischen Unterstützungsbedarfs erfolgt im schriftlichen Gutachten die Festlegung der Kategorie (Unterstützungsbedarf ja/nein) sowie Empfehlungen zum Modell der Schule, zur Bewertung, zu den nächsten Förderzielen, zu den Methoden der Förderung und in einem inklusiven System auch ein Anspruch auf wirksame Unterstützung.

In einem pädagogisch-diagnostischen Gutachten sind die drei Schritte der pädagogischen Diagnostik getrennt voneinander vorzufinden. Die drei Schritte beziehen sich auf eine Diagnostik zu einem Zeitpunkt und ist daher eine Statusdiagnostik. Als Prozessdiagnostik sind die drei Schritte adaptierbar, wenn die Diagnostik wiederkehrend erfolgt.

Arbeitsauftrag:
Beschreiben Sie eine kritische Situation im Praktikum möglichst deskriptiv und ohne Bewertungen. Welche Bewertungen der Situation sind möglich? Beschreiben Sie mehrere Alternativen.

Take-Home-Message:
Vorschnelle Bewertungen und Festlegungen sind kein Zeichen eines professionellen Vorgehens. Wertneutrale Beschreibungen, deskriptive Darstellungen und einen umfassenden Pool an Informationen kurz darzustellen ist dagegen der Kern im Schreiben und Kommunizieren der pädagogischen Diagnostik.

1.5 Von der Person zu den Umweltbedingungen

Traditionell steht im Mittelpunkt der pädagogischen Diagnostik das Kind selbst als Einzelfall im Fokus. Meist wird Bezug auf die Frage genommen, welche Schule, Fördergruppe oder Institution (früher Anstalt) die passende wäre. Die individuellen Dispositionen des Kindes werden dabei als stabile Eigenschaften (Traits) betrachtet (kognitive Fähigkeiten, (schulische) Kompetenzen, Verhaltensweisen und Motivation). Die Betrachtung der Person selbst ist notwendig, um ein individuelles Profil der Person zu erhalten. Mittlerweile wird dabei nicht ausschließlich eine medizinisch-individuelle Sichtweise für die pädagogische Diagnostik verwenden, sondern eine soziale Sichtweise, welche den Einfluss, Wirkungsweise und die Anforderungen der Umwelt und der pädagogischen Institutionen miteinbezieht, um eine Aussage über die zukünftige Entwicklung zu treffen.

Ab dem Ende des 19. Jahrhunderts hatte die Psychopathologie zunehmend einen Einfluss auf die Schule (Bühler & Hofmann, 2017). Hierbei wurde mittels verschiedener Störungsbilder untersucht, ob das dementsprechende Kind tauglich für die Schule war oder eine besondere Form der Förderung benötigte. Auch der erste Intelligenztest von Binet und Simon wurde 1905 konstruiert, um neben dem Lehrer:innenurteil auch einen objektiveren Vergleich bei der Überweisung in eine Sonderschule zu haben (Binet & Simon,1912). Der Test unterschied anhand von 30 verschieden abgestuften Aufgaben zwischen dem als normal eingestuften Kind und einem Kind mit Schwachsinn. Nach Bühler (2017) wollte Binet erreichen, dass jedes Kind die passende Hilfe bekommt anstelle von Strafen, Schelten und Prügel, welcher sich die Lehrkräfte bei fehlendem Lernfortschritt bedienten. Hierbei bezieht sich das pädagogische Handeln auf eine Diagnostik nach medizinischem Vorbild:

„Die Pädagogik ist mit der Medizin zu vergleichen, sie enthält zugleich eine Diagnostik und eine praktische Behandlung des Schülers. … und nach Auffinden einer augenscheinlichen Ursache wird man zur Erforschung ihrer bestmöglichen Gegenmittel schreiten" (Binet, 1912, S. 34 f.).

Die Lehrkräfte sahen nach Binet (1912) Kinder mit niedrigen Leistungen als faul an, welche einen schwachen „Willen" (Binet, 1912, S. 240) hatten. Hier setzte der erste Intelligenztest an und zeigte, dass Kinder über verschiedene kognitive Fähigkeiten verfügten, welche nach dem medizinisch-individuellen Modell von Behinderung, auch eine andere Art von Unterricht oder Schule implizierten, wie beispielsweise die Didaktik der Vorsorge (Klauer, 1975). Nach dieser wird Lernen behavioristisch gesehen wird und Kinder mit Lernbehinderung mehr Zeit, kleinschrittiger und strukturierteren Unterricht benötigen (Klauer, 1975).

Weitere Informationen:
In Kapitel 2 und 3 bei Bundschuh & Winkler (2019): Einführung in die sonderpädagogische Diagnostik. Ebenso enthält Tröster (2019) einen vertieften Einblick in die Messung und Bewertung von Intelligenz.

Auch heutzutage wird in der Praxis unter Diagnostik insbesondere die Messung von individualen Dispositionen am Kind, seinen Eigenschaften, Fertigkeiten und Kompetenzen verstanden. Hierbei werden nun auch soziale Ursachen und die Interaktion mit dem Umfeld als Ursachen angesehen und nicht mehr rein die Seele oder Veranlagung als Ursachen angenommen. Ebenso wird anerkannt, dass institutionelle und gesellschaftliche Barrieren eine Rolle im diagnostischen Prozess spielen. Dies hat sich auch in den internationalen Klassifikationen niedergeschlagen. So wird mittlerweile nicht das Defizit der betroffenen Person in den Vordergrund gestellt, sondern das Problem als Wechselwirkung zwischen den individuellen Dispositionen der Person und der Umwelt gesehen. Während die International Classification of Impairments, Disabilities and Handicaps (WHO, 1980) Beeinträchtigung (Impairment) noch rein individuell-medizinisch als Defizit in der Person definiert hat, sieht die International Classification of Functioning, Disability and Health (WHO, 2001) Behinderung nun als Wechselwirkung zwischen Person und Umwelt (Body Functioning, Participation, Activity).

Solche Klassifikationen sind nach aktueller Rechtslage notwendig, um einer Person im Schul- oder Gesundheitssystem zusätzliche Hilfe und Unterstützung zukommen zu lassen. Hierbei muss die Problematik der Person (Beeinträchtigung) meist nachweislich und messbar abweichend von der Normalität liegen, also außerhalb der Norm. So wird nach dem individuell-medizinischen Modell eine Störung oder, wenn umfänglich und langandauernd, eine Beeinträchtigung festgestellt. Diese Form der Diagnostik ist im Gesundheitssystem etabliert und wird dort mit Hilfe von Klassifizierungssystemen wie mittlerweile der ICD 11 von der Weltgesundheitsorganisation (WHO) beschrieben. So finden sich dort unter 06 Neurodevelopmental Disorders folgende Störungsbilder, welche von Kinder- und Jugendpsychiater:innen diagnostiziert werden und auch im schulischen Kontext bekannt sind:

- 6A00 Disorders of Intellectual Development
- 6A01 Developmental Speech or Language Disorders
- 6A01.0 Developmental Speech Sound Disorder
- 6A01.1 Developmental Speech Fluency Disorder
- 6A01.2 Developmental Language Disorder
- 6A02 Autism Spectrum Disorder
- 6A03 Developmental Learning Disorder
- 6A04 Developmental Motor Coordination Disorder
- 6A05 Attention Deficit Hyperactivity Disorder

Um die Diagnose eines solchen Störungsbildes zu rechtfertigen, müssen die beschriebenen Symptome langfristig und an mehreren Orten bzw. Situationen auftreten (länger als sechs Monate) sowie die Person im Alltag behindern. Nach den Leitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF, 2018) ist es beispielsweise für die Diagnostik einer Rechenstörung notwendig, in einem normierten Schulleistungstest Mathematik ein Ergebnis mit einer Differenz zum Mittelwert der Alters- oder Klassennorm von mindestens 1,5 Standardabweichungen (d. h. PR ≤ 7) zu erreichen. Gibt es noch weitere qualitative und klinische Kriterien, so kann die Differenz auch nur eine Standardabweichung betragen. So wird mittels eines psychometrisch geprüften und normierten Tests ein wesentlicher Abstand zur Altersgruppe festgestellt und mit den weiteren Kriterien der Diagnostik abgewogen. Hierbei ist ebenso wichtig, dass die Symptome nicht von anderen Störungen oder fehlender Pädagogik verursacht sind.

Arbeitsauftrag:
Vergleichen Sie die Beschreibungen von 6A03 developmental learning disorder und 6A05 Attention deficit hyperactivity disorder. Welche Bedarfe und Unterstützungsformen ergeben sich aus den Beschreibungen? Inwieweit hilft Ihnen diese Beschreibung für die pädagogische Arbeit?

Exkurs:
Ärzt:innen verwenden Leitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) zur Diagnostik und Therapie von Störungen. Sie finden diese Leitlinien frei zugänglich unter: https://www.awmf.org/leitlinien

Diese Kategorien und Beschreibungen der ICD 11 werden im Gesundheitssystem angewendet. Für die Schule haben sie keine unmittelbare Bedeutung, da die ICD 11 nicht in Schulgesetzen oder Verordnungen implementiert ist. Diagnosen und Besonderheiten werden in ärztlichen Gutachten beschrieben und können von der Schule berücksichtigt werden. Meist ist das ärztliche Gutachten dann Teil eines schulpsychologischen oder sonderpädagogischen Gutachtens, welches konkrete Interventionen und Möglichkeiten, wie beispielsweise den Nachteilsausgleich oder den sonderpädagogischen Unterstützungsbedarf des Schulsystems, beschreibt.

Exkurs:
Die Verwendung von Störungsbildern wird in der Pädagogik generell kritisch gesehen, da diese als Label neben einem Nutzen auch einen Schaden erzeugen. Neben anerkannten medizinischen Störungsbildern gibt es eine Reihe von weiteren Störungsbildern, welche in der Öffentlichkeit diskutiert werden. Besonders umstritten war und ist hier beispielsweise die Minimal Brain Dysfunction (Clemens, 1966), welche angeblich insbesondere bei Kindern mit Lernschwierigkeiten als neuronale Störung auftritt. Diese Störung ist jedoch zu minimal, um messtechnisch nachweisbar zu sein. Obwohl es zu dieser These hinter dem Störungsbild kaum belege gibt, ist es in Literatur und Praxis immer noch anzutreffen.

Take Home Message:
Insbesondere kritisch ist eine unreflektierte Übernahme von Störungsbilder in der Schule, wenn diese Störungsbilder als Rechtfertigung remediale Förderkonzepte nach einem individuell-medizinischen Modell von Behinderung benutzt werden.

1.6 Ebenen der Diagnostik

Die pädagogische Diagnostik umfasst Faktoren im Kind, im sozialen Umfeld und im schulischen Umfeld, welche je nach Fragestellung und pädagogischem Modell unterschiedlich gewichtet werden. Für die Pädagog:innen ist hierbei auch relevant, welche Prozesse sie anregen können und wie eine Veränderung erzielt werden kann. Eine Verhaltensänderung beim Kind selbst zu erzielen ist möglich, aber mitunter sehr aufwendig. Ebenso ist auch das soziale Umfeld des Kindes meist als stabil zu betrachten und Veränderungen sind nur langsam möglich. Das pädagogische oder schulische Umfeld und die Lernumgebung dagegen kann die Lehrkraft kontrollieren und steuern. Daher muss eine pädagogische Diagnostik neben dem Kind auch die pädagogische Situation und die Gestaltung der Lernumgebung umfassen.

1.6.1 Diagnostik auf der Ebene der Person

Auf der Ebene des Kindes werden seine individuellen Dispositionen, Fähigkeiten und Kompetenzen erhoben. Für den Bereich des schulischen Lernens sind dies beispielsweise folgende Aspekte, welche Lutz und Gebhardt für ihr Raster verwendet haben:

- Schulleistung, Kompetenzen, Vorwissen
- Lern- und Arbeitsverhalten, (Lern-)Motivation
- Soziale und emotionale Kompetenzen
- Sprachliche Kompetenzen, Kommunikation
- Wahrnehmungsmöglichkeiten und motorische Kompetenzen
- Kognitive Fähigkeiten, Metakognition und Lernstrategien

Take-Home-Message:
Die alleinige Messung von kindzentrierten Daten wie Intelligenz reicht nicht für eine Interpretation aus. Eine Bewertung benötigt immer auch Daten des Umfeldes und des pädagogischen Kontexts.

1.6.2 Diagnostik auf der Ebene des häuslichen und schulischen Umfelds

Für die Entstehung von Lernschwierigkeiten sind die aktuelle pädagogische Situation und das soziale sowie schulische Umfeld ebenso entscheidend wie die individuellen Dispositionen der lernenden Person. Der Mensch interagiert mit seinem Umfeld. So entstehen Probleme bis hin zur Behinderung aus der Wechselbeziehung mit der Umwelt. Jede Person hat eigene Stärken und Schwächen, welche je nach Kontext unterschiedlich wirken. Meist sind es der Kontext, das Vorwissen und die Strategie, welche über eine erfolgreiche Handlung und das Überwinden von Barrieren entscheiden. Ebenso können die Barrieren durch die Gestaltung der Lehrkraft selbst abgebaut und gemindert werden. Sind Probleme in der Schule aufgetreten, können die Probleme mit den Eltern gelöst werden. So können im positiven Fall Eltern und das soziale Umfeld wie ein Schutzschirm für das Kind wirken und es findet ein gemeinsames konstruktives Arbeiten zwischen Elternhaus und Schule statt.

Die Schule selbst hat je nach Schulart unterschiedlich hohe **Ansprüche und Anforderungen** an die Schüler:innen. Diese könnten beispielsweise sein: Kinder sollen pünktlich, ausgeschlafen, satt und sauber in die Schule kommen. In ihrer Schultasche sind die ordentlich erledigten Hausaufgaben, das Federmäppchen ist vollständig mit gespitzten Stiften. Nach der Schule haben die Kinder die Möglichkeit an einem Arbeitsplatz in Ruhe ihre Hausaufgaben zu machen und bekommen von ihren Eltern Hilfe. Das Kind kann selbständig die Hausaufgaben erledigen, lernt mit und wiederholt vor Prüfungen. Ein Computer und Internet steht den Kindern daheim zur Verfügung und deren Benutzung wird von Eltern unterstützt sowie überwacht. Diese Aufzählung der Anforderung ist je nach Lehrkraft und Schule unterschiedlich. Je nach finanziellen und bildungsrelevanten Möglichkeiten der Familie sind manche der genannten Punkte sind für einige Kinder mit ihren Familien kaum zu erfüllen

und nur für wenige Familien möglich. Daher legt auch die Schule mit ihren Ansprüchen und Unterstützungsleistungen fest, wann Problemlagen entstehen. Mittlerweile bieten einige Schulen vor der Schule eine Frühbetreuung mit Frühstück an, welches von der Kommune unterstützt wird, damit kein Kind hungrig in die Schule muss.

Kinder scheitern in der jeweiligen Lernsituation, wenn sie keinen Zugang zum Inhalt bekommen, nicht schnell genug verstehen oder mit der Situation an sich überfordert sind. Wird diese Situation nicht durch die Lehrkraft, das Kind selbst oder das Umfeld gelöst, wirkt sich dies auf die Motivation und die Selbstwirksamkeit bezogen dem schulischen Fach aus. Manifestiert sich die Schwirigkeit kann ein Teufelskreis Lernstörung (Betz & Breuninger, 1998) beginnen und zu Missmotivation und Schulunlust führen. Gründe für schulisches Scheitern sind vielfältig und **Unterrichtsbeobachtungen** und eine **Analyse der Lernumgebung** und des bereitgestellten Materials sowie eine Analyse der bearbeiteten Aufgaben und der Lösungswege der Person sind notwendig. Handlungsleitend ist hierbei die Frage, ob es sich nur um eine kurzfristige Lernschwierigkeit handelt, welche sich mit anderen pädagogischen Handlungen wie beispielsweise anderen Erklärungen, anderem Material oder einer vertieften Übung verbessern lässt. Oder aber es besteht keine adäquate Passung zwischen den schulischen Anforderungen und der lernenden Person und die akademische und soziale Entwicklung der lernenden Person ist gefährdet. Daraus folgt die Frage, warum die Passung nicht gegeben ist und ob im Elternhaus oder im schulischen Umfeld eine andere Form der Unterstützung bereitgestellt werden kann oder sollte.
Schulprobleme lassen sich somit meist nicht direkt kausal mit dem häuslichen Umfeld erklären, aber es besteht ein Zusammenhang. Dieser ist zwar nicht so stark wie Korrelationen auf der individuellen Ebene des Kindes in Bezug auf die akademischen Kompetenzen, aber er wird in allen Large Scale Assessments wie PIRLS und PISA gemessen. Daher haben Kinder aus sozial benachteiligten Familien überrepräsentativ häufiger als andere Kinder einen sonderpädagogischen Unterstützungsbedarf im Lernen oder Verhalten (Bildungsbericht, 2014; Gebhardt et al., 2015b). Für Kinder aus sozial benachteiligten Familien ist es schwieriger, die explizit und vor allem die implizit gesetzten Ansprüche der Schule zu erfüllen. Umso selektiver das Schulsystem ist, umso mehr Schwierigkeiten haben solche Familien, da sie weniger Unterstützung in den Bereichen Ausstattung, Lernen, Nachhilfe usw. geben können. Ebenso muss die Frage nach irgendeiner **Form einer besonderen Belastung** für das Kind in der Diagnostik abgeklärt werden, welche die schulische Entwicklung gefährden kann. (Im Fallbuch von Lutz und Gebhardt (2021) sind solche Fälle beschrieben.)

Daher gehört die Erfassung des häuslichen und schulischen Umfelds zur Anamnese in der pädagogischen Diagnostik. Meist sind die dahinterliegenden Risiko- und Unterstützungsfaktoren im Einzelfall in gemeinsamer Arbeit mit der Schulsozialarbeit beispielsweise mittels standardisierter Fragebögen wie des Einschätzungsbogens zum sozialen und häuslichen Umfeld von Schulkindern (SHU-P) (Lutz & Gebhardt, 2022) zu klären und gegebenenfalls Lösungen anzuleiten. Das Vorgehen hierzu beruht auf dem ökosystemischen Ansatz von Bronfenbrenner (1979), welcher vom Kind ausgehend mehrere verzahnte Systeme sieht.

Take-Home-Message:
Das soziale Umfeld einer Person ist relativ stabil und kann extern nur bedingt beeinflusst werden. Daher konzentriert sich die Lehrkraft auf die Umwelt der Bildungsinstitution selbst, welche sie aktiv gestalten kann.

Eine weitere systemische Frage ist es, ob sich nicht die Schule an die Bedürfnisse des Kindes anpassen kann, um Inklusion bei Kindern mit sonderpädagogischen Unterstützungsbedarf zu ermöglichen. So hat die **Kind-Umfeld-Analyse** als Diagnostik das Ziel Barrieren für die Inklusion eines individuellen Kindes in der Schule zu erfassen und wenn möglich abzubauen. Der Ansatz nach Hildeschmidt und Sander (1993) stammt aus der inklusiven Pädagogik. Es priorisiert die Förderdiagnostik unter Einbezug des Umfeldes und

der Eltern statt einer Platzierungsdiagnostik und wirbt für die Anpassung der allgemeinbildenden Schule als Ort der Inklusion. Sander (1998) gibt zu dem Konzept folgende Grundlagen:

- Nicht das Kind, sondern das Kind-Umfeld steht im Mittelpunkt.
- Die kindgegebenen Umstände der Schule müssen mitberücksichtigt werden.
- Die schulische Veränderung ist das Ziel der Diagnostik.
- Die Diagnostik erfolgt im Team.
- Es erfolgt in zeitlichen Abständen eine Wiederholung der Diagnostik.
- Die subjektive Sichtweise der Beteiligten muss in die Diagnostik einbezogen und reflektiert werden.
- Im Schulleistungsbereich müssen verschiedene Bewertungsmaßstäbe mitberücksichtigt werden neben dem klassenbezogenen und lehrplanbezogenen Vergleich auch die individuelle Entwicklung.

Mehr Informationen:
Heimlich, U. (2022). Kind-Umfeld-Analyse und ökologische Modelle.

Auf der Ebene des häuslichen Umfeldes haben Lutz und Gebhardt für ihr Raster folgende Aspekte näher beleuchtet:

Erziehungsberechtigte – Familie

- Familiärer Hintergrund
- Unterstützende Maßnahmen und Institutionen
- Lernorganisation daheim
- Freizeitverhalten
- Kontakt zur Schule

Externe Partner:innen

- Außerschulische Förderung
- Therapien und Behandlungen

Auf der Ebene des schulischen Umfeldes für den Bereich des schulischen Lernens sind dies beispielsweise folgende Aspekte:

Schule und Unterricht

- Lernumgebung, Classroom Management
- Lehrer:innen-Schüler:innen-Interaktion
- Individuelle (Förder-) Maßnahmen, Nachteilsausgleich, Notenschutz/-befreiung
- Unterstützende Partner:innen/Institutionen

Peers in der Schule

- Kooperatives Lernen
- Schüler:innen-Schüler:innen-Interaktion

Video:
Bronfenbrenners ökologisches System - 5 Kräfte die unser Leben beeinflussen (7:14)

Weitere Informationen:
Heimlich, U. (2022). Kind-Umfeld-Analyse und ökologische Modelle. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung*. *Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 17-24). Universität Regensburg.

Version 0.1: Soziales und häusliches Umfeld von Schulkindern: Eine Einschätzung für PädagogInnen (SHU-P)

Arbeitsauftrag:
Wie heißen die Ebenen nach Bronfenbrenners Modell und warum sind diese Ebenen für die schulische Intervention nach Heimlich wichtig?

1.6.3 Das Einzelfallraster für die pädagogische Diagnostik

Eine Möglichkeit, die vorher besprochenen Ebenen pädagogischen Diagnostik zu ordnen, ist das Einzelraster von Lutz (2023). Dieses Einzelfallraster ist im Fallbuch von Lutz & Gebhardt (2023) näher erläutert.

Mehr Informationen:

Eine PDF-Vorlage für das Raster zum Ausdrucken finden sie auf der Homepage des Lehrstuhls unter folgendem Link: Raster Version 01

1.7 Glossar – Übersicht über Begriffe im Bereich Diagnostik

Die Methoden und Instrumente der Diagnostik sind verschieden. Nachfolgend werden diese Begriffe skizziert, um einen Überblick zu geben:

Lernen ist ein Prozess, welcher aktiv von den Lernenden gesteuert wird und Wissen konstruiert.

Lerntransfer geht über einen Lern- oder Übungseffekt hinaus. Es kann positiven oder negativen Lerntransfer geben (Klauer, 2011).

Lehrziel beinhaltet, welcher Inhalt und Kompetenzbereich mit welchen Lehr-Lernprozessen und -aktivitäten erreicht werden soll (Klauer & Leutner, 2012).

Statusdiagnostik ist eine meist umfangreiche Diagnostik mit mehreren Instrumenten (Beobachtungen und Tests) zu einem Messzeitpunkt zur Messung eines Ist-Standes im Lernen und Verhalten.

Prozessdiagnostik ist eine Diagnostik mit mehreren Messungen mit einem oder verschiedenen Instrumenten über einen längeren Zeitraum. Förderdiagnostik sollte daher immer auch eine Prozessdiagnostik sein.

(Lern-)Verlaufsdiagnostik wird mit einem Instrument zu mehreren Messungen verwendet, um einen Verlauf möglichst reliabel zu dokumentieren.

Lernprozessdiagnostik bezieht sich ebenso auf den Lernstand, die Förderung und den Lernverlauf. Meist wird aber neben quantitativen Daten insbesondere mittels qualitativer Daten ein Verlauf dargestellt.

Informationen können **qualitativ** (als Text) und/oder **quantitativ** (als Zahl) erhoben werden. Informationen können dabei **standardisiert** erhoben werden, um möglichst vergleichbar zu sein oder **nicht-standardisiert**, um möglichst spezifisch die Beobachtung zu beschreiben.

Bei Tests gibt es **informelle Verfahren**, welche nur theoretisch konstruiert sind oder Verfahren, welche psychometrisch nach den Gütekriterien geprüft und meist auch normiert sind. Testverfahren können je nach **Bezugsnorm** unterschiedlich konstruiert sein. So gibt es die individuelle Bezugsnorm für die Lernentwicklung eines Kindes, die soziale

Bezugsnorm als Vergleich mit allen Kindern und die kriteriale Bezugsnorm als Vergleich mit einem externen Referenzrahmen.

Weitere Informationen zu den Bezugsnormen:
Jungjohann, J. & Gebhardt, M. (2022). Bezugsnormorientierung im Unterricht. In. M. Gebhardt, M. Schurig & D. Scheer (Hrsg.). *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 25-32). Universität Regensburg.

Pädagogische Diagnostik besteht aus einer systematischen Sammlung und Interpretation pädagogisch relevanter Daten, welche einerseits aus standardisierten, mittels psychometrisch geprüfter Tests und Fragebögen erhobenen Daten und andererseits auch aus unstandardisierten qualitativ beschreibenden Daten bestehen. Beide Formen von Daten werden benötigt. Eine Diagnostik ohne standardisierte Daten ist nicht vergleichbar und kaum nachzuvollziehen und es besteht die Gefahr von Willkür. Eine Diagnostik ohne beschreibende, auf die Person und ihr soziales und persönliches Umfeld angepasste Beschreibung kategorisiert nur grob vereinfachend und kann wesentliche Aspekte der pädagogischen Fragestellung übersehen.

Instruktionspsychologie (Instructional design) beschäftigt sich mit der Instruktionsplanung oder Instruktionsgestaltung in der Planung und Gestaltung des Lehr-Lern-Prozesses und des Curriculums (Klauer & Leutner, 2012).

(Lern-)Therapie ist eine außerschulische, strukturierte, umfangreiche, evidenzbasierte Förderung in mehreren verzahnten (Einzel-)Sitzungen, um eine diagnostizierte Lernschwierigkeit oder Störung einer Person individuell zu kompensieren, abzumildern oder zu beseitigen. Aktuell verfügt das Schulsystem in Deutschland nicht über solche Möglichkeiten. Daher finden im Unterricht mit sonderpädagogischer Unterstützung, wenn überhaupt, nur einzelne therapeutische Methoden oder Maßnahmen statt. Integrative Lerntherapie nach Bender et al. (2017) ist eine Form der Lerntherapie, welche durch einen eigenen Fachverband vertreten wird.

Das **Multi-Tiered Systems of Support (MTSS)** genannte **Mehrebenenpräventionsmodell** bündelt Fördermaßnahmen, meist gestuft auf drei Ebenen, in einem neuen, formalen, aber zusammenhängenden System (siehe Kapitel im Einführungswerk).

2 Statusdiagnostik oder Verlaufsdiagnostik

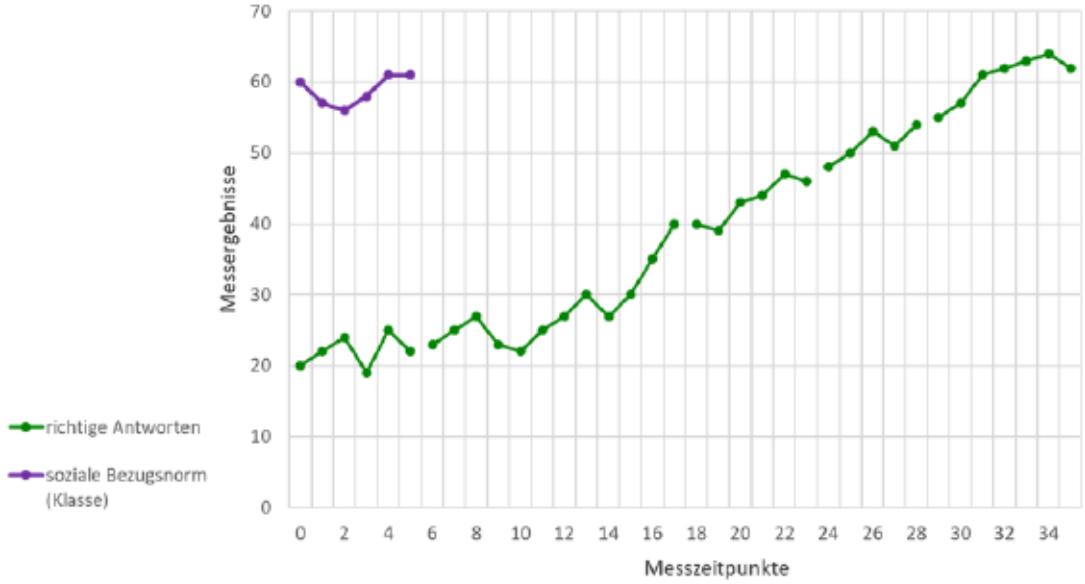
In der Statusdiagnostik haben diagnostische Verfahren eine lange Tradition, die zu einem Zeitpunkt den Ist-Stand von mehreren Teilkompetenzen (z. B. Facetten von Schulleistungen, Sprachentwicklung oder Intelligenz) erheben. Häufig werden hierbei psychologische Tests verwendet, welche normiert und standardisiert schulische Kompetenzen meist multidimensional in ihrer gesamten Breite zu einem Zeitpunkt messen. So wird der Leistungsstand der Lernenden in den getesteten Dimensionen zum Testzeitpunkt erfasst. Die Einzel- oder Gruppentests haben normierte Vergleichswerte in Bezug auf eine repräsentative Vergleichsgruppe. So wird für die Identifikation oder Klassifikation einer gravierenden Schwierigkeit ein Ergebnis unter einem Prozentrang 16 (1. Standardabweichung) zu den Ergebnissen einer repräsentativen Erhebung aller Schüler:innen verwendet. Zusätzlich können informelle Methoden und Beobachtungen zur Erfassung der Lernausgangslage und des aktuellen Status eingesetzt werden.

Statusdiagnostikverfahren haben den Vorteil, dass sie die Erstellung eines sehr differenzierten Profils von Lernenden ermöglichen. In diesen Profilen werden Stärken und Schwächen dokumentiert. Die Profile geben Anhaltspunkte für die Förderziele und für die Entwicklung des Förderplans (Heimlich et al., 2021). Die psychologischen Schulleistungstests haben Alters- und Klassenstufennormen, welche für eine Bewertung der Leistung einer oder eines Lernenden mit einer Vergleichsgruppe benötigt werden. Der Vergleich mit einer spezifischen Vergleichsgruppe dient dazu, abweichende Leistungen festzustellen. Statusdiagnostische Verfahren werden zur Erstellung von Gutachten, wie dem Gutachten zur Feststellung eines Bedarfs an sonderpädagogischer Unterstützung oder zur Wahl des Förderortes, benötigt (siehe Kapitel 3). Ein Nachteil der Statusdiagnostik ist, dass Durchführung und Auswertung für die Lehrkräfte sowie für die Schüler:innen meist als zeitintensiv beschrieben werden. Daher lohnt sich der Einsatz vorwiegend, wenn ein Überblick über die gesamte Breite der Kompetenzen benötigt wird oder weitreichende Entscheidungen getroffen werden. Die meisten Statustests sind für einen Messzeitpunkt ausgelegt und nicht zur wiederholten Messung geeignet, daher können nur wenige Verfahren zur Evaluation von Förderungen eingesetzt werden.

Verlaufsdiagnostische Verfahren (Voß & Gebhardt, 2017) messen mittels mehrerer Erhebungen den Fortschritt in den Bereichen Lernen oder Verhalten und melden die Ergebnisse unmittelbar meist in Form eines Graphen an die Lehrkräfte zurück. Der Verlauf wird reliabel und fair gemessen, um während der Förderung den Lehrkräften den Erfolg oder eben auch Nicht-Erfolg einer Maßnahme rückzumelden. Daher wird diese Form der Diagnostik auch formative Diagnostik genannt. In Abbildung 2 ist ein exemplarischer Lernverlauf anhand eines einminütigen Lesetests dargestellt. Grüne Punkte stehen für die Anzahl der richtig gelösten Wörter zum jeweiligen Testzeitpunkt, während die lila Punkte die kumulierten Ergebnisse der Klassenkamerad:innen darstellen.

Take-Home-Message:
Während Statusverfahren einen Messzeitpunkt umfassen, benötigen verlaufsdiagnostische Verfahren mehrere Messzeitpunkte mit den gleichen oder vergleichbaren Instrumenten.

Abbildung 2
Ergebnisgraph einer Lernverlaufsdiagnostik



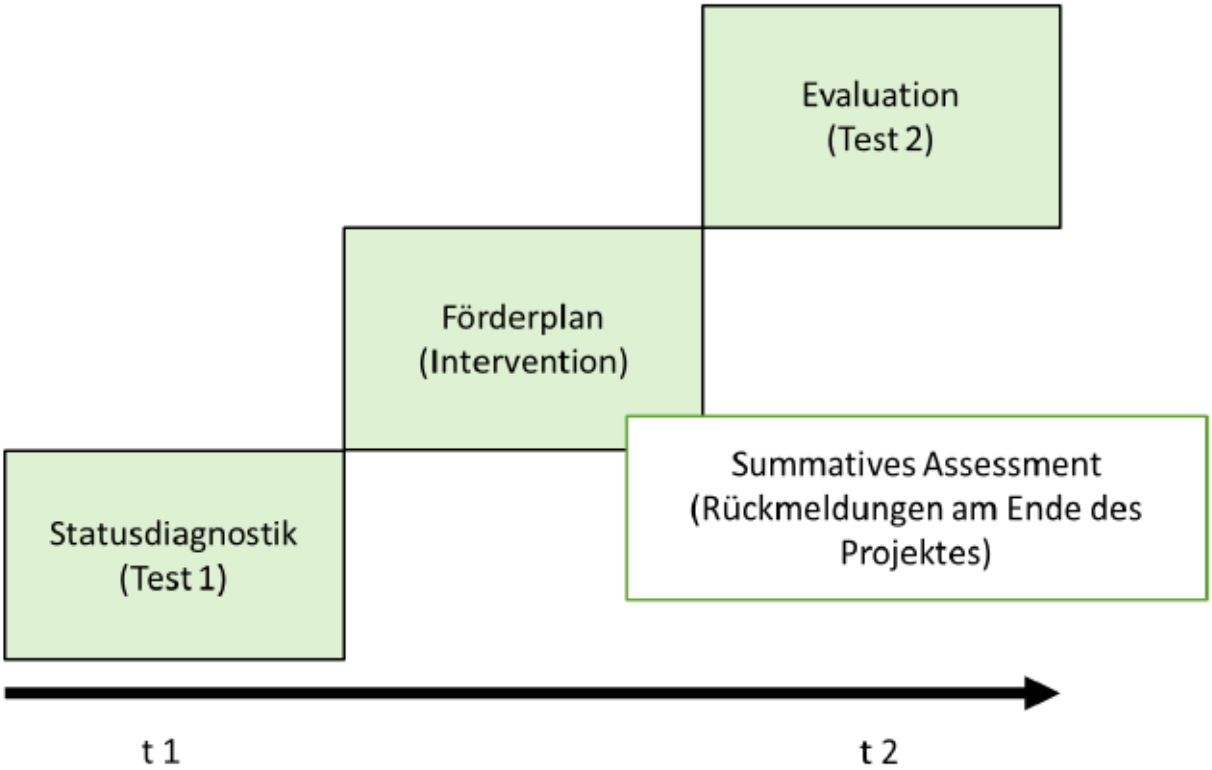
Arbeitsauftrag:
Beschreiben Sie in eigenen Worten die Unterschiede zwischen einer Status- und einer Verlaufsdiagnostik. Was sind die jeweiligen Anwendungsformen?

2.1 Modelle der Prozess- oder Förderdiagnostik

Im Rahmen der pädagogischen Diagnostik wird die aktuelle Ausgangslage beschrieben und es werden mögliche Bedarfe beim Kind und der schulischen Umgebung begründet. Die Frage ist, welche Ressourcen benötigt werden, um welche Ziele in welcher Zeit zu erreichen. Für die pädagogische Diagnostik steht daher nicht allein das Kind im Vordergrund, sondern die Empfehlung, Planung und Durchführung von pädagogischen Maßnahmen.

DIE Modell von Heimlich
Heimlich (2012) sieht die drei Schritte aus Diagnostik, Intervention und Evaluation (DIE) als ein Prozessmodell der pädagogischen Förderung, welches in Abbildung 3 dargestellt ist. Eine Diagnostik in der Schule hat den Vorteil, dass alle Teilprozesse des Modells in der Schule organisiert werden können und daher auch Informationen darüber vorliegen. Somit wird aus einer Statusdiagnostik eine Prozessdiagnostik, welche einen Prä- und Posttest umfasst und auch den Erfolg der Maßnahmen nachweisen kann. Erstaunlicherweise wird das Modell aber selten mit zwei Erhebungen in der schulischen Praxis umgesetzt. Häufiger anzutreffen ist in der pädagogischen Praxis eine einmalige Statusdiagnostik mit Förderplanung, welche bei Schüler:innen mit sonderpädagogischem Unterstützungsbedarf vorgeschrieben ist. Da die Art und Form der Evaluation nicht festgelegt sind, wird diese meist nicht in der genügenden Form des DIE Modells (Heimlich, 2012) praktiziert und dokumentiert.

Abbildung 3
Das DIE Modell von Heimlich in vereinfachter Form



Arbeitsauftrag:
Beschreiben Sie das DIE Modell in eigenen Worten. Wird dieses Modell in der schulischen Praxis aktuell angewendet?

Modell von Vernooij
Ähnlich dem DIE Modell von Heimlich sieht auch Vernooij (2013) die sonderpädagogische Diagnostik als regelmäßige Statusdiagnostik zur Klärung des Ist-Standes und zur Evaluation der Förderung. Die Prozessdiagnostik ist eine Erweiterung mit mehreren diagnostischen Schritten:

„Als Prozessdiagnostik wird eine lernwegbegleitende Diagnostik bezeichnet, bei der in regelmäßigen Abständen der aktuelle Ist-Stand des zu fördernden Kindes überprüft wird, sowohl hinsichtlich des Leistungsstandes in den schulischen Fächern als auch hinsichtlich des Arbeits- und Sozialverhaltens (je nach Förderschwerpunkt unterschiedlich gewichtet). Das heißt, Prozessdiagnostik ist die flexible bzw. variable individuums- und bedarfsorientierte Anwendung diagnostischer Verfahren (z. B. Schulleistungstest) und Methoden (z. B. Beobachtung, Dokumentenanalyse, Fehleranalyse) über einen längeren (Förder-)Zeitraum mit dem Ziel, die Effektivität der Förderung zu überprüfen und Leistungs- und Verhaltensveränderungen des Kindes festzustellen." (Vernooij, 2013, S. 34)

Eine Diagnostik mit verbundener Förderplanung, Durchführung der Förderung und Evaluation wird insbesondere im sonderpädagogischen Feld als Förderdiagnostik bezeichnet. Die Durchführung der Förderung steht hierbei im Vordergrund, während Fragen zur Klassifikation und Selektion in den Hintergrund rücken. Der Begriff wurde eingeführt, da die Diagnostik in der schulischen Praxis häufig als sogenannte Platzierungsdiagnostik oder Feststellungsdiagnostik (KMK, 2019) für den sonderpädagogischen Unterstützungsbedarf durchgeführt wurde, um einen Schulort zuzuweisen oder vom Schulsystem mehr Ressourcen und Unterstützung zu bekommen. Es besteht daher immer die Gefahr, insbesondere in remedialen pädagogischen Systemen, dass die eigentliche Klassifikation die eigentliche

pädagogische Fragestellung der Diagnostik zu stark beeinflusst (Gebhardt, 2021). Dies kann so weit gehen, dass der eigentlich helfende Erkenntnisgewinn im pädagogischen Prozess nicht mehr gesucht und beachtet wird und die jeweiligen Ergebnisse so beeinflusst werden, dass eine bestimmte Klassifikation und systemische Zuweisung zustandekommt.

Take-Home-Message:
Die Modelle der traditionellen Förderdiagnostik setzen sich aus den Phasen Diagnostik (Messung 1), Intervention (Förderung) und Evaluation (Messung 2) zusammen.

Arbeitsauftrag:
Beschreiben Sie das Modell von Vernooij (2013) in eigenen Worten. Wo gibt es Unterschiede in den jeweiligen Sichtweisen auf den diagnostischen Prozess.

Weitere Informationen:
Der Artikel von Moser Opitz (2022) vergleicht die unterschiedlichen Ansätze und die Entwicklung der Diagnostik über die Zeit in der Fachdidaktik der Mathematik und der Sonderpädagogik.

GRID Modell
Eine Arbeitsgruppe der Universität Köln entwickelte das Gesamtübergreifende Richtungsweisende Individuelle Diagnostik (GRID) – Modell (Hammes-Schmitz et al., 2021). Anhand von förderdiagnostischen Lernbeobachtung wird ein Raster ausgewertet. Ziel ist es mehrere für den Schulalltag wichtige Dimensionen für die Lehrkraft einfach zu visualisieren, zeitökonomisch anwendbar zu sein und einen ressourcenorientierten Blick zu fördern.

Internetseite des GRID-Modells

2.2 Modell zur Lernverlaufsdiagnostik

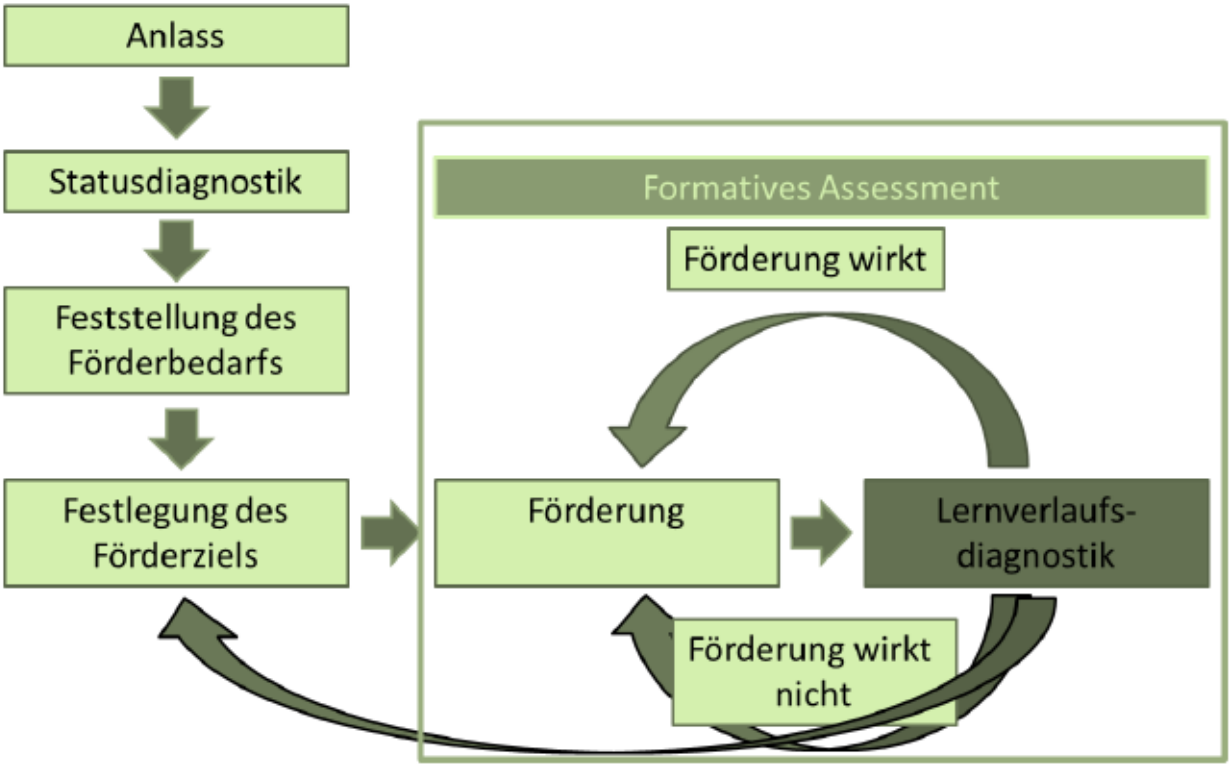
Mit Verlaufsdiagnostik wird der Frage nachgegangen, ob der aktuelle Unterricht oder die Förderung auf das Lernziel wirken oder nicht (Fuchs & Fuchs, 2006). Ziel ist es, Lernschwierigkeiten zu vermeiden und mit möglichst präventiven Methoden eine möglichst positive Schulkarriere zu erreichen.

Verlaufsdiagnostik zielt daher besonders auf Kinder mit schulischen Schwierigkeiten ab, bei welchen die Gefahr besteht, dass der reguläre Unterricht oder die angewandte Methode nur bedingt helfen. Mittels Verlaufsdiagnostik kann frühzeitig eine Nicht-Passung erkannt werden und die Methodik daraufhin angepasst werden. Verlaufsdiagnostik muss daher sensibel die Verläufe von Kindern mit schulischen Schwierigkeiten erfassen und ist immer bei einem längerfristigen Kompetenzaufbau sinnvoll. Die Lernverlaufsdiagnostik ist ein Test, welcher zu mehreren Testzeitpunkten in drei bis fünf Minuten durchgeführt werden kann und die Lernentwicklung reliabel und sensibel misst (siehe für ein Review zum Anfangsunterricht im Lesen Jungjohann et al. (2018)). Im Gegensatz zu den im letzten Kapitel dargestellten Modellen der Förderdiagnostik wird der Prozess nicht anhand unterschiedlicher Instrumente oder verschieden breiter Statusdiagnostiken grob eingeschätzt, sondern anhand eines Instrumentes reliabel gemessen (siehe Abbildung 4).

Während die meisten Modelle der Prozessdiagnostik wie beispielsweise das DIE Modell (Heimlich et al., 2021) aus drei Schritten besteht, verbindet die Verlaufsdiagnostik die Schritte der Förderphase und der Evaluation durch ein **formatives Assessment**. Formativ bedeutet hier, dass die diagnostische Information des Tests direkt für die Förderung rückgemeldet und genutzt werden kann. Dies ist ein zentraler Bestandteil des RTI Modells.

In Abbildung 4 wird das Prozessmodell der pädagogischen Diagnostik mit Verlaufsdiagnostik (Gebhardt & Jungjohann, 2020; adaptiert von Wilbert, 2014) in zwei Bereichen dargestellt. Der vertikal angelegte Bereich ist die Statusdiagnostik (siehe Abbildung 4, linker Abschnitt). Nach einem Anlass und einer Formulierung einer pädagogischen Fragestellung werden zu einem Messzeitpunkt quantitative und qualitativ relevante und notwendige Daten zur Person, dem Umfeld und der pädagogischen Situation erhoben. Diese Daten werden gemeinsam im Team besprochen, mögliche Ursachen und pädagogische Bedürfnisse ermittelt und daraus Ziele, Prognosen und nötige Unterstützungshilfen festgelegt. Eine umfängliche Statuserhebung und eine Festsetzung von Entwicklungszielen sollten in größeren zeitlichen Abständen (halbjährlich/jährlich) wiederkehrend erfolgen (Hartung et al., 2021). Eine solche Diagnostik ist zeitlich und personell umfänglich und aufwendig. Dafür ermöglicht diese Diagnostik ein umfassenderes Bild, um pädagogische Handlungsweisen zu prüfen und Ziele festzulegen.

Abbildung 4
Prozessmodell zur Lernverlaufsdiagnostik



Der Verlauf der Linie vom Förderziel über die Förderung zur Verlaufsdiagnostik und zurück stellt in Abbildung 4 den Kreislauf dar, welcher durch das unmittelbare Feedback der formativen Verlaufsdiagnostik möglich ist. Förderungen und Veränderungen der Lernumgebung wirken meist nicht für alle Personen gleich und benötigen eine gewisse Zeit, bis Erfolge sichtbar sind. Durch die vielen kurzen Messzeitpunkte wird ein genereller Trend ermittelbar sein, wenn das Ziel und das verlaufsdiagnostische Verfahren sensibel genug gewählt wurden. Ist ein solcher Trend nicht erkennbar, kann die Lehrkraft bereits während der Förderung Übungen ändern, intensivieren oder, wenn die erwarteten Maßnahmen nicht helfen, dass ursprüngliche Ziel noch einmal überprüfen.

Take-Home-Message:
Förderungen mit Lernverlaufsdiagnostik umfassen mehrere formative Messungen, welche direkt auf die Förderung wirken sollen.

Die Wirksamkeit von Verlaufsdiagnostik und formativen Verfahren im Generellen liegen an ihrem unmittelbaren Feedback und der Dokumentation. Ohne dieses Feedback besteht die Gefahr, dass die Förderung wie vorher geplant ohne Änderung durchgeführt worden wäre und das Ergebnis erst am Ende in der summativen Evaluation gesehen wird. Ebenso gilt, dass selbst wenn die Lehrkraft die Förderung anhand ihrer Beobachtungen abgebrochen oder verändert hätte, sie beim summativen Vorgehen keine Dokumentation über die Effektivität der Förderung an sich hat. Daher lohnt sich Verlaufsdiagnostik insbesondere bei Kindern mit schulischen Schwierigkeiten, bei denen eine längere und intensivere notwendige Förderung geplant ist.

Video:
Feststelldiagnostik, Förderdiagnostik und Lernverlaufsdiagnostik (16:22)

Arbeitsauftrag:
Ist ein Prozessmodell zur LVD immer ein Kreislauf?

Weitere Informationen:
Gebhardt, M. & Jungjohann, J. (2020). Analyse der Lernausgangslage und der Lernentwicklung - Prozesse der Förderdiagnostik. In U. Heimlich & F. Wember (Hrsg.), *Didaktik des Unterrichts bei Lernschwierigkeiten: Ein Handbuch für Studium und Praxis* (S. 367–380). Kohlhammer.

Ebenbeck, N., Jungjohann, J. & Gebhardt, M. (2022) Inklusive Lernverlaufsdiagnostik mit der Onlineplattform Levumi. Eine Übersicht für die Praxis. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 783–791). Universität Regensburg.

2.3 Zone der nächsten Entwicklung

Lew Wygotski (1896 – 1934) war einer der ersten bis heute bekannten Forscher, welcher Menschen mit Behinderung nach einem sozialen Modell betrachtete. Seine Vorstellung über die kognitive Entwicklung und das Denken und Sprechen (2002) sowie sein soziales Verständnis von Behinderung waren der damaligen Zeit weit voraus. Diese Gedanken hatten einen starken Einfluss auf spätere Forscher:innen, welche in der sogenannten kulturhistorische Schule zusammengefasst werden. Seine Gedanken zur Entwicklung sind nach wie vor aktuell und führten unter anderem auch zum Ansatz des dynamischen Testens (Dynamic Assessment), bei dem es nicht um den aktuellen Status, sondern das mögliche Potential von Personen geht.

Wygotski sah die soziale Interaktion des Individuums als wesentlich an. Für die pädagogische Diagnostik ist bedeutend, dass er zwischen der **Zone der aktuellen Entwicklung** und der **Zone der nächsten Entwicklung** unterschied. Ein Kind kann in der Zone der aktuellen Entwicklung bestimmte Handlungen allein durchführen. Wenn ein Kind in dieser aktuellen Zone Hilfestellung erhält, kann es weitere Handlungen erlernen. Dieser Bereich des Möglichen ist die Zone der nächsten Entwicklung, die das Kind aktuell noch nicht erreicht, aber mit Hilfestellung und Unterstützung erreichen kann, um in dieser nächsten Zone allein zu handeln. Der Blick von Pädagog:innen sollte nach diesem Ansatz somit nicht allein auf die aktuelle Phase und einer Messung des Status Quo liegen, sondern auch auf der zukünftige Entwicklung und den nächsten Schritten des Lernens.

Take-Home-Message:
Auch in der aktuellen Praxis wird meist zu sehr auf den aktuellen Status geachtet und zu wenig die Zone der nächsten Entwicklung in den Blick genommen. Insbesondere für die Kompetenzentwicklung ist diese Zone aber die entscheidende für pädagogische Handlungen. Somit erkannte Wygotski früh die hohe Bedeutung von Prozessdiagnostik für das Lernen.

Video von Sprouts zu Wygotskys Theorie der kognitiven Entwicklung

Mehr Informationen zum dynamischen Testen:
Das dynamische Testen (Guthke, 1996) beinhaltet die Idee, systematische Hilfestellungen während der Testung zu geben und die Reaktion der Personen auf die Hilfestellung als Teil des zu messenden Konstrukts zu sehen.

Gabriel, T., Börnert-Ringleb, M. & Wilbert, J. (2022). Dynamisches Testen im Spannungsfeld von Selektion und Modifikation. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 313-321). Universität Regensburg.

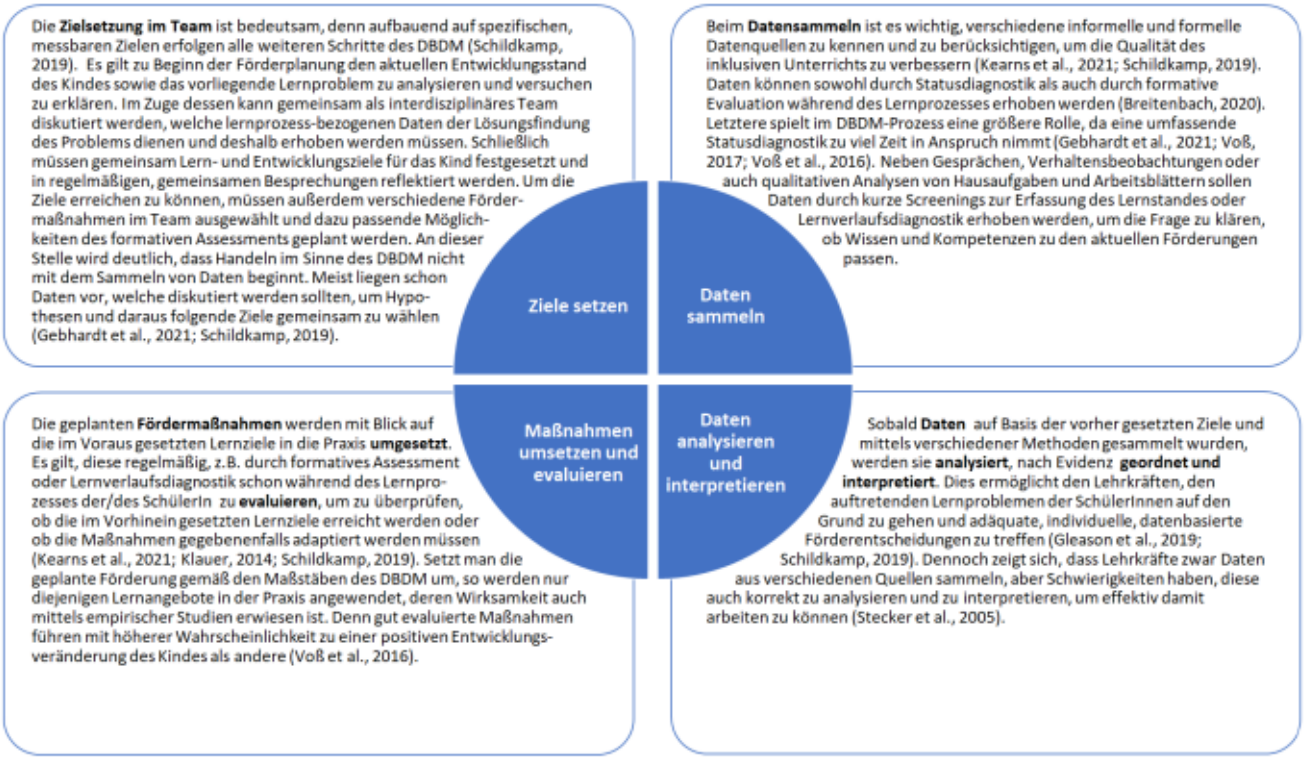
2.4 Data Based Decision Making

Data Based Decision Making (DBDM) ist ein neuer Name für die pädagogische diagnostische Entscheidung. Ähnlich wie Klauer (1978) generell pädagogische Diagnostik als Erkenntnisbemühung sieht, geht es um die richtige Entscheidung für pädagogische Handlungen. Im Unterschied zu Klauer (1978) wird in diesem Konzept der Prozess betont und sich auf die Frage konzentriert, welche Informationen als Daten vorliegen und für die weitere Entwicklung des Kindes wichtig sind. Da ein Kind in der Schule über mehrere Jahre lernt und arbeitet, entstehen im Lernprozess Daten und es ist möglich, standardisierte Instrumente wiederkehrend zu verwenden. DBDM ist somit eine wiederkehrende **Prozessdiagnostik** mit dem Ziel individuelle Unterstützung zu bestimmen, zu realisieren und zu evaluieren. Neben der Statusdiagnostik stehen somit die formative Evaluation der Lernentwicklung bei gleichzeitiger Förderung im Vordergrund (Voß et al., 2016). Das Konzept des DBDM bezieht sich auf den Response-to-Intervention-Ansatz (RTI) (Brown-Chidsey & Steege, 2010; Espin et al., 2017; Voß et al., 2016). Insbesondere wenn sich die datenbasierte Entscheidung auf Unterricht oder Lernen bezieht, wird für dieses Konzept auch der Name Instructional Decision Making (Hamilton et al., 2009) verwendet.

DBDM beinhaltet die Bausteine (siehe Abbildung 5):

- Ziele setzen
- Daten sammeln, analysieren und interpretieren
- Maßnahmen umsetzen und evaluieren
- Daten analysieren und interpretieren

Abbildung 5
Data Based Decision Making nach Lutz et al. (2022)



Artikel:
Lutz, S., Boschner, S. & Gebhardt, M. (2022). Data-based Decision Making (DBDM) in der inklusiven Diagnostik und Förderplanung. In. M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 33-42). Universität Regensburg.

Arbeitsauftrag:
Bitte lesen Sie den Text über Data Based Decision Making (DBDM) in der inklusiven Diagnostik und Förderplanung von Lutz, Boschner und Gebhardt (2022).
Beschreiben Sie die Problemlage von Frau Schmidt und die Umsetzung von DBDM durch Frau Schmidt in eigenen Worten. Gibt es alternative Lösungen für Frau Schmidt?

Was ist der Unterschied und die Gemeinsamkeiten von DBDM zu den regulären Modellen der Förderdiagnostik, wie beispielsweise dem DIE Modell von Heimlich?

Take-Home-Message:
Data Based Decision Making (DBDM) ist ein Teamprozess für Prozessdiagnostik, welcher nur gemeinsam in der Schule implementiert werden kann.

3 Sonderpädagogische Diagnostik

Sonderpädagogische Diagnostik wendet die Methoden und Instrumente der pädagogischen Diagnostik auf sonderpädagogische Fragestellungen an. Meist geht es um das Identifizieren und Klassifizieren von Behinderung und Beeinträchtigung sowie die Unterstützung und Förderung von Menschen mit Behinderung. Je nach Lebensabschnitt und Bereich basieren diese Feststellungen und Förderungen entweder auf den Schulgesetzen und Verordnungen der jeweiligen Bundesländer oder für die Zeit vor und nach der Schule auf den jeweiligen Sozialgesetzbüchern. Ebenso gibt es verschiedene sonderpädagogische Unterstützungsbedarfe sowie Behinderungen. Auf die jeweiligen Unterschiede in den Bundesländern (Wolf & Dietze, 2022) und die einzelnen sonderpädagogischen Schwerpunkte sei an dieser Stelle auf die einzelnen Kapitel im Herausgeberwerk Handbuch sonderpädagogische Diagnostik (2022) verwiesen.

Kapitel aus dem Herausgeberwerk:
Wolf, L. M. & Dietze, T. (2022). Ein Überblick über die Organisation der Feststellung von sonderpädagogischen Förderbedarfen in Deutschland. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 325-344). Universität Regensburg.

Hartung, N., Vossen, A., Hecht, T. & Sinner, D. (2022). Diagnostik zur Feststellung des sonderpädagogischen Unterstützungsbedarf im Schwerpunkt Lernen. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 517-526). Universität Regensburg.

Stattdessen wird auf die Gemeinsamkeiten der sonderpädagogischen Diagnostik und der pädagogischen Diagnostik näher eingegangen. Sonderpädagogische Diagnostik ist Teil der inklusiven Schule und des Schulsystems. Die Feststellung und Förderung im Bereich der Sonderpädagogik betrifft je nach Bundesland 7 bis 10% aller Schüler:innen (Statistik der KMK, 2019). Die pädagogischen Entscheidungen, welche mit sonderpädagogischer Diagnostik verbunden sind, sind für die Beteiligten und auch für das Schulsystem gravierend und häufig kostenintensiv. Das Thema sonderpädagogische Diagnostik ist daher kein Randthema der pädagogischen Diagnostik, sondern die eigentliche Königsdisziplin von Diagnostik in der Schule. Die Problemlagen sind meist komplex und Lösungen durch die Beteiligten und auch durch die meist verschiedenen Unterstützungssysteme nicht immer einfach.

3.1 Ziele der sonderpädagogischen Diagnostik

Sonderpädagogische Diagnostik war und ist mit der Zuweisung zu einer bestimmten Schule oder Förderung verbunden. Der **sonderpädagogische Förder- bzw. Unterstützungsbedarf (SPF/SPU) ist daher ein schulspezifisches systemisches Konstrukt.** Die Vergabe des Unterstützungsbedarfs wird benötigt, um einem einzelnen Kind sonderpädagogische Hilfe und weitere Unterstützung zu ermöglichen. Nach den Definitionen in den jeweiligen Gesetzen wird definiert, dass sonderpädagogische Unterstützung dann erfolgt, wenn die reguläre Unterstützung des Kindes durch die allgemeine Schule nicht ausreicht, um die soziale und schulische Entwicklung des Kindes zu fördern. Um einen Zugang zu den Bildungsinhalten, Teilhabe und Chancengleichheit zu ermöglichen, wird dann ein sonderpädagogischer Unterstützungsbedarf vergeben. Dieser ist in den jeweiligen Gesetzen der Bundesländer definiert und wird mit einem sonderpädagogischen Gutachten begründet.

Bereits mit der Gründung der ersten Hilfsschulen wurde für die Erkennung oder Zuweisung der „richtigen" Kinder eine Diagnostik verwendet. Diese war in den Anfangsjahren von Konzepten aus der Medizin geprägt (Siehe Kapitel 1.6.1, um beispielsweise die Diagnostik

und Kategorien für geistige Schwäche festzustellen. So ist die Diagnostik insbesondere in den Händen der Schulmediziner:innen, wie es Garz (2022) für Berliner Schulen von 1845 und 1914 beschreibt.
Schulärzt:innen sind auch nach 1945 noch Teil der sonderpädagogischen Diagnostik und gehören nach der KMK (1960) zum sonderpädagogischen Überprüfungsverfahren. In diesem erstellt die/der Schularzt/-ärztin ein Gutachten, um zu klären, ob „Zusammenhänge zwischen dem physischen und psychischen Zustand des Kindes und seinem Versagen in der Volksschule bestehen" (KMK, 1960, S.29). Mittlerweile wird die Frage nach den Zusammenhängen der schulischen Probleme mittels Tests zur Intelligenz und den kognitiven Fähigkeiten von Psycholog:innen, Schulpsycholog:innen oder Sonderpädagog:innen geklärt. Medizinische Gutachten spielen zu einzelnen Fragen von Beeinträchtigungen und Behinderung eine Rolle, sind aber nicht mehr automatisch Teil einer sonderpädagogischen Diagnostik.

Take Home Message:
Die Diagnostik hat sich von einer personen- und eigenschaftsorientierten Diagnostik zu einer verhaltens- /leistungsbezogenen und situationsbezogenen Diagnostik verändert.

Ein Kind erhält eine sonderpädagogische Unterstützung in der inklusiven Schule oder in der Förderschule, wenn es mittels einer **„Feststellungsdiagnostik"** (KMK, 2019, S. 11), meist durchgeführt von einer sonderpädagogischen Lehrkraft, untersucht wurde und ein sonderpädagogisches Gutachten mit Hinweisen zu Umfang und Art der Unterstützung zugewiesen bekommen hat (Gebhardt & Jungjohann, 2020). Die Grundlage für eine zusätzliche Ressource sowie das zugrundeliegende Vergabeverfahren sind in den Gesetzen und Verordnungen der einzelnen Bundesländer festgeschrieben. Zwar unterscheiden sich die einzelnen Bundesländer in ihren Gesetzen, in der Umsetzung der inklusiven Beschulung und auch in der sonderpädagogischen Förderquote, jedoch gibt es auch Gemeinsamkeiten in den einzelnen deutschen Bundesländern (Sälzer et al., 2015) sowie hinsichtlich der Umsetzung in Österreich (Buchner & Gebhardt, 2011) und der Schweiz (Mejeh & Powell, 2018). Die Vergaben des sonderpädagogischen Unterstützungsbedarfs werden über das Schulamt von allen Beteiligten auf Grundlage einer Diagnostik gemeinsam entschieden und ausgehandelt. Die Entscheidung fällt im Einzelfall, bei dem auf der einen Seite die Bedürfnisse des Kindes und auf der anderen Seite die Unterstützungsmöglichkeiten der Schule betrachtet werden. Während in der Vergangenheit hier nur die Auswahl des Schulortes möglich war, besteht die Hoffnung, dass in Zukunft auch mehrere Unterstützungssysteme und Möglichkeiten in der inklusiven Schule für besondere pädagogische Bedarfe berücksichtigt werden.

Feststellungsdiagnostik = oder ≠ Selektionsdiagnostik?
In einem Schulsystem mit eigenen Förderschulen oder eigenen Sonderklassen, besteht immer die Gefahr, dass eine Feststellungsdiagnostik nicht nur anhand des individuellen sonderpädagogischen Förderbedarfs des Kindes gefällt wird, sondern von einer Vielzahl an systemischen Gründen beeinflusst wird. Je nachdem wie viele Kinder die allgemeinbildende Schule aufnehmen oder abgeben möchte, kann die ursprüngliche Konzeption und die eigentliche Feststellung eines sonderpädagogischen Förderbedarfs verzerrt werden.

Eine **Feststellungsdiagnostik** als reine Klassifikation der einzelnen Förderschwerpunkte bzw. Beeinträchtigungen wird als unscharf kritisiert und hilft den beteiligten Personen nicht weiter. Daher muss eine sonderpädagogische Diagnostik auch immer den Aspekt der Unterstützung in der Schule sowie konkrete, adäquate und evidenzbasierte Förderempfehlungen umfassen. Aus diesem Grund umfasst eine sonderpädagogische Diagnostik neben der Feststellungsdiagnostik (Klassifikation) auch immer eine **Förderdiagnostik** (KMK, 2019), welche die Festsetzung von Lernzielen, die Planung von individuellen Bildungsangeboten und im optimalen Fall auch deren Evaluationsmerkmale umfasst. Die Förderdiagnostik ist eine Prozessdiagnostik und erweitert die Statusdiagnostik mit Intervention und

Evaluation oder auch mit Verlaufsdiagnostik. Im Sinne der inklusiven Pädagogik ermöglicht die Förderdiagnostik auch eine Änderung der Schule und eine Veränderung der schulischen Förderung, um eine bessere Passung zwischen dem Kind und dem schulischen Umfeld zu erreichen.

Diese Idee basiert auf der Kind-Umfeld-Analyse nach Hildenschmidt und Sander (1993). Förderdiagnostik wird als Ermöglichung der Umsetzung der schulischen Inklusion unter Einbezug des Umfeldes und der Eltern gesehen. Sander (1998) gibt zu dem Konzept folgende Grundlagen (Siehe dazu vertieft Heimlich, 2022):

- Nicht das Kind, sondern das Kind-Umfeld steht im Mittelpunkt.
- Die kindgegebenen Umstände der Schule müssen mitberücksichtigt werden.
- Die schulische Veränderung ist das Ziel der Diagnostik.
- Die Diagnostik erfolgt im Team.
- Es erfolgt in zeitlichen Abständen eine Wiederholung der Diagnostik.
- Die subjektive Sichtweise der Beteiligten muss in die Diagnostik einbezogen und reflektiert werden.
- Im Schulleistungsbereich müssen verschiedene Bewertungsmaßstäbe mitberücksichtigt werden – neben dem klassenbezogenen und lehrplanbezogenen Vergleich auch die individuelle Entwicklung.

Video: Kind Umfeld Analyse (14:47)

Der Begriff Förderdiagnostik wird von unterschiedlichen Autor:innen und Konzepten verwendet. Daher gibt es verschiedene Auslegungen und Meinungen, welche Schritte sowie Instrumente eine Förderdiagnostik umfasst. Eine begriffliche Unklarheit ist hierbei, ob Förderdiagnostik auch die Förderung miteinschließt. Das Wort Diagnostik selbst bezieht sich auf den Erkenntnisprozess für eine pädagogische Handlung. In einem nächsten Schritt kann diese Handlung diagnostisch evaluiert werden. Die pädagogische Handlung selbst, wie eine Förderung, ist dabei nicht Teil der Diagnostik. Die Intervention ist als eigenständig zu sehen. Die Intervention wird anhand der theoretischen und evidenzbasierten Lösung der Problemlage und anhand der Ergebnisse aus der Diagnostik ausgewählt, abgeleitet und begründet.

„Diagnostik kann nicht fördern, Diagnostik kann feststellen, was ist - und auch nur in Grenzen, warum das so ist, wie es ist" (Klauer, 2005, S. 191).

Während in der wissenschaftlichen Literatur die sonderpädagogische Förderung als wesentlicher Teil der (Förder-)Diagnostik gesehen wird, besteht in der Praxis die Gefahr, dass die Diagnostik als Feststellungsdiagnostik in Form der Klassifikation den Ansprüchen genügt. Moser Opitz et al. (2019) befragten 226 Lehrkräfte in der Schweiz zum Einsatz von Förderplänen in inklusiven und separativen Klassen der Primar- und Sekundarstufe. Die Erstellung von Förderplänen wurde von den Befragten als eine bürokratische Notwendigkeit für eine Dokumentations- und Legitimationsfunktion gesehen, während der Aspekt der Förderung als weniger wichtig eingeschätzt wurde. Da sich sonderpädagogische Arbeit an der Effektivität der Förderung beurteilen lassen muss, sind solche Ergebnisse erschütternd.

Weitere Leseempfehlungen:
Förderdiagnostik als Lernprozessdiagnostik nach Eberwein (1995)
Interview mit Alfred Sander von Frank R. Müller (2018)

Weitere Informationen:
Im Herausgeberwerk Förderdiagnostik von Kornmann, Meister & Schlee (1994) werden die verschiedenen Konzepte und Sichtweisen auf Förderdiagnostik diskutiert:
Kornmann, R. Meister, H. & Schlee, J. (1994). *Förderdiagnostik. Konzepte und Realisierungsmöglichkeiten*. Universitätsverlag C. Winter.

Buchrezension von Klauer (2005) zum Buch Neue Entwicklungen in der Förderdiagnostik von Mutzeck & Jogschies (2004) (Paywall-Lesbar mit VPN an der Universität)

3.2 Erstellung eines sonderpädagogischen Gutachtens

Um eine Verbindung zu den im Einzelfallraster gesammelten Informationen herzustellen, werden Teile des Einzelfallrasters von Lutz und Gebhardt (2022) in das Gutachten eingebunden.

Ausgehend von einer pädagogischen Fragestellung aus einer schulischen Beratungssituation wird ein mögliches Vorgehen, adaptiert nach dem Leitfaden nach Heimlich et al. (2021), als Fördergutachten dargestellt. Dieses wird als Fördergutachten bezeichnet, da die nachfolgenden Empfehlungen für die Förderung expliziter Bestandteil des Gutachtens sind. Es geht über viele gesetzliche Vorschriften hinaus, welche keine oder kaum pädagogische Handlungsempfehlungen verlangen.1. Untersuchungsanlass und Fragestellung

2. Vorgeschichte und Klärung der pädagogischen Situation
   a. Außerschulisches Umfeld
   b. Vorschulisches bzw. schulisches Umfeld
3. Hypothesen inkl. Untersuchungsplanung und Auswahl der diagnostischen Verfahren
4. Erhebung und Darstellung der Untersuchungsergebnisse
5. Interpretation der Untersuchungsergebnisse
6. Stellungnahme – Zusammenfassende Beschreibung des individuellen sonderpädagogischen Förderbedarfs
7. Fördermaßnahmen und Ressourcen – Empfehlungen für pädagogisches Handeln und Vorhersage
8. Entscheidung aus den pädagogischen Gesprächen und weitere pädagogische Empfehlungen

Arbeitsauftrag:
Warum benötigt man eine pädagogische Fragestellung und was wird unter einer Hypothese verstanden?

Lesen sie das Fallbeispiel Noah im Fallbuch von Lutz & Gebhardt (2023) ab Seite 204. Ist der Fall anhand des Rasters als sonderpädagogisches Gutachten nachvollziehbar dargestellt? Welche weiterführenden pädagogischen Entscheidungen und Handlungen würden sie für ein inklusives Schulsystem wünschen und erwarten?

Untersuchungsanlass & Fragestellung
Kontaktaufnahme mit den Eltern bzw. den Lehrkräften. Was ist die konkrete Problemstellung? Welche Person nimmt welches Problem wahr? In welchem pädagogischen Setting befindet sich die Person?
Methode: Gespräche
➔ Erste Eingrenzung der Problemstellung und Bildung möglicher Fragestellung

Vorgeschichte und Klärung der pädagogischen Situation
Wie ist die Entwicklung des Kindes seit der Geburt verlaufen? Welche Bildungsinstitutionen hat das Kind besucht? Wurden besondere Förder- bzw. Therapiemaßnahmen durchgeführt? Liegen bereits diagnostische Befunde vor? Welche Unterstützungsmaßnahmen gibt es in der Familie und im sozialen Umfeld? Wie ist die aktuelle Lernsituation?
Methode: Gespräche, Beobachtung & Sichtung der Schülerakten sowie weiterer Nachweise
➔ Weitere Eingrenzung der Problemstellung und Bildung der Fragestellung

Hypothesen inkl. Untersuchungsplanung und Auswahl der diagnostischen Verfahren
Welche Annahmen und Argumente liegen dem Fall zugrunde? Was sind mögliche Ursachen für die Problemlage? Aus der Ausgangssituation werden die Fragestellung formuliert und Hypothesen gebildet. Welche Informationen werden auf Ebene des Kindes, des Umfeldes und der Schule benötigt?
Auswahl und Begründung der Instrumente:

- Beobachtungen
- Analyse von Schülerdokumenten (Stand der Kompetenz, Strategien, Fehleranalyse)
- Gespräche
- Standardisierte und informelle Tests

➔ Entwicklung des Untersuchungsplans

Erhebung und Darstellung der Untersuchungsergebnisse
Gab es Besonderheiten oder Störungen bei der Beobachtung oder bei der Durchführung von Tests? Wie wurden die Untersuchungsverfahren in der Diagnosesituation eingesetzt? Gibt es sonstige Faktoren, die das Ergebnis im Untersuchungsverlauf möglicherweise beeinflusst haben?

➔ Ziel: Prüfung der Objektivität der Untersuchung

Welche Ergebnisse haben die einzelnen Verfahren erbracht?
Wie können die Ergebnisse einzelnen Verhaltens- und Lernbereichen zugeordnet werden?

➔ Ziel: Übersichtliche Darstellung der Ergebnisse

Interpretation der Untersuchungsergebnisse
Zusammenfassung des gemessenen Ist-Standes und der pädagogischen Situation. Bestätigen die Ergebnisse die Fragestellung? Welches Profil innerhalb der Person wurde gefunden? Welche Stärken und welche Schwierigkeiten liegen vor? Wie sind die Ergebnisse der Person im Vergleich zur Altersgruppe?
Welche Hinweise auf die Änderung der pädagogischen Situation und welche möglichen Fördermaßnahmen lassen sich ableiten?

➔ Ziel: Herausarbeitung der Grundlagen für die pädagogische Entscheidung

Stellungnahme – Zusammenfassende Beschreibung des individuellen sonderpädagogischen Förderbedarfs
Welche sonderpädagogischen Bedürfnisse liegen nach den gesetzlichen Verordnungen vor? Welche Klassifikationen können wie begründet werden?

Fördermaßnahmen und Ressourcen – Empfehlungen für pädagogisches Handeln und Vorhersage
Welche pädagogischen Ziele und weiteren Schritte kann man aus den Ergebnissen ableiten? Welche Entwicklungen sind unter den gegebenen Umständen möglich? Welche Entwicklung ist mit weiteren Maßnahmen möglich? Wie kann man mit welchen Methoden die Ziele erreichen?
Wie sollten diese Fördermaßnahmen didaktisch-methodisch realisiert werden?
Welche Ressourcen, pädagogischen Modelle und Schulformen stehen für die Realisierung einer besseren pädagogischen Situation zur Verfügung?

➔ Ziel: Grundlagen für ein Elterngespräch mit allen Beteiligten

Entscheidung aus den pädagogischen Gesprächen und weitere pädagogische Empfehlungen
Die Entscheidung über die Klassifikation und mögliche Maßnahmen kann nur gemeinsam mit den Eltern getroffen werden. Aus diesem Grund sind die Elternbeteiligung und eine

freie Entscheidung der Eltern zwischen verschiedenen Alternativen die Voraussetzung für ein gesetzlich verankertes freies Elternwahlrecht.
Den Eltern werden der aktuelle Ist-Stand, die bisherige Entwicklung und die Möglichkeiten der aktuellen Schule (aktuellen Förderung) vorgestellt. Ebenso werden weitere Möglichkeiten der Förderung und alternative Förderorte und die damit verbundene Entwicklung dargestellt. Hierbei sollten nach gesetzlichen Vorgaben und empirischen Erkenntnissen inklusive Fördersysteme bevorzugt werden. In der Praxis ist dies nur bedingt der Fall. Die Eltern entscheiden sich meist für das für sie passende Angebot. Dabei haben die Eltern selten gleichwertige Auswahlmöglichkeiten. Die Eltern müssen einzelne Aspekte wie beispielsweise Teilhabe, soziale Partizipation, individuelle Förderung, Ganztagsbetreuung, wohnortnahe Schule oder therapeutische Angebote gewichten und ihre Entscheidung treffen.

➔ Ziel: Entscheidungsfindung

Weitere Informationen:
Vossen, A., Hartung, N., Hecht, T. & Sinner, D. (2022) Das sonderpädagogische Gutachten (Status und Feststellungsdiagnostik). In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 345-352). Universität Regensburg.

Arbeitsaufträge:
Lesen Sie den Beitrag von Vossen et al. (2022). Was ist nach dieser Quelle ein sonderpädagogisches Gutachten? Welche Kriterien sind für ein gutes Gutachten relevant?

Welche Gemeinsamkeiten und Unterschiede gibt es zwischen dem hier im Buch vorliegenden Konzept und dem Beitrag von Vossen et al. (2022)?

Take-Home-Message:
Ein Gutachten kann je nach Fragestellung und Auftraggeber unterschiedlich gestaltet sein. Ziel ist es die pädagogische Fragestellung nachvollziehbar für Außenstehende zu beantworten und das gewählte Vorgehen zu dokumentieren und zu begründen.

3.4 Förderplan

Ein sonderpädagogischer Förderplan, im Englischen Individual Educational Plan (IEP) genannt, ist für alle Kinder mit festgestelltem sonderpädagogischen Förderbedarf verpflichtend. Er beschreibt auf individueller Ebene die Förderziele, die pädagogischen Maßnahmen und den angestrebten Verlauf sowie deren Evaluation (siehe die Zusammenfassung von Prince et al. (2018) unter Kapitel 1.1). Der Förderplan (IEP) besteht aus den folgenden Bestandteilen:

1. Der Förderplan beinhaltet die **aktuelle Situation, den Lernstand** und die aktuelle mögliche Entwicklung oder den Verlauf ohne weitere Förderung. Der Förderplan fasst die zentralen Ergebnisse der vorher stattgefundenen Diagnostik zusammen. Dies können sowohl Ergebnisse einer umfassenden Statusdiagnostik als auch einer Prozessdiagnostik sein, welche als Einzelfallraster dargestellt werden.

2. Der Förderplan beschreibt den bzw. die **Förderbereich/e und die einzelnen Förderziele**, welche mit einem Konzept und einzelnen Maßnahmen erreicht werden können. Das Förderziel ist angemessen gewählt und führt zu einer Verbesserung der persönlichen Situation des Kindes. Das Erreichen der Förderziele kann kriterial nachgewiesen werden. Das Förderkonzept mit den Maßnahmen ist theoretisch sowie in Bezug zur Evidenz begründbar und praktisch unter den gegebenen

Umständen vor Ort umsetzbar. Die gewählten Maßnahmen sind potenziell wirksam und Fortschritte unter Bezug auf die Förderung und Veränderung der schulischen Umgebung möglich. Bei der Wahl des Förderziels ist bereits darauf zu achten, wie dieses später überprüft und evaluiert werden soll.

3. Der Förderplan stellt die Ressourcen der Schule, eine mögliche Veränderung der Schule und die zusätzlichen pädagogischen **Maßnahmen im Einzelnen** dar. Hier kann auch auf die vorliegenden Informationen zurückgegriffen werden, die im Einzelfallraster gesammelt wurden. Welche Personen sind in die Förderplanung miteinbezogen? Wo und wann finden pädagogische Maßnahmen statt und welche interne oder externe Unterstützung sind dafür notwendig? Welche Maßnahmen finden innerhalb des Unterrichts (unterrichtsparallele Unterstützungsangebote), im Einzelfall (additive Angebote außerhalb des Unterrichts), in Kleingruppen oder außerhalb der Schule, beispielsweise in Förderzentren, statt? Wie können eine höchstmögliche inklusive Förderung und volle Teilhabe gewährt werden? Wie können stigmatisierende Tendenzen vermieden werden?

4. Der Förderplan wird **summativ oder formativ evaluiert** und muss halbjährlich fortgeschrieben werden. Daher folgt auf den aktuellen Förderplan meist der nächsten Förderplan. Eine Messung, wie und ob die Ziele erreicht wurden, ist immer Teil des Förderplans. In den USA wird hierbei eine Verlaufsdiagnostik gefordert, da nur mit Tests mit mehreren Erhebungen auch ein Effekt für die Maßnahme abgeleitet werden kann (Prince et al., 2018). Ebenso ist es möglich, kriteriale Tests und Beschreibungen zum Erreichen des Förderziels zu verwenden, welche für Außenstehende nachvollziehbar sind.

Die Förderpläne unterscheiden sich je nach den gesetzlichen Anforderungen und den Anforderungen der Schuladministration. Ebenso sind die Förderpläne nach der pädagogischen Maßnahme und den Konzepten je nach dem sonderpädagogischen Schwerpunkt verschieden konzipiert. Darüber hinaus gibt es je nach dem pädagogischen Modell unterschiedliche Namen und Schwerpunktsetzungen in Inklusion und Förderschule.

Arbeitsauftrag:
Beschreiben Sie in wenigen Worten, was ein Förderplan ist und welche Elemente dieser beinhalten soll.
Vergleichen Sie im nächsten Praktikum die Förderpläne der Schüler:innen mit einem sonderpädagogischen Unterstützungsbedarf.

Weitere Informationen:
Bernasconi, T. (2022) ICF-orientierte Förderplanung. *In* M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 733-747). Universität Regensburg.

Jungjohann, J. & Hüninghake, R. (2022). Förderplanung im Bereich Rechtschreibung. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 697-707). Universität Regensburg.

4 Messen und Bewerten im pädagogischen Kontext

Bewertungen und Rückmeldungen sind Teil jeder Bildungsinstitution und auch Teil der damit verbundenen pädagogischen Diagnostik. Bereits im 17. Jahrhundert wurden Schüler:innen von den unterrichtenden Priestern, Mönchen und Schwestern meist anhand ihres Fleißes und ihrer Frömmigkeit bewertet. Im Laufe der Zeit wechselte das Bewertungsraster von einer rein religiösen Bezugsnorm hin zu einem pädagogischen Ansatz der damaligen Zeit (Berdelmann, 2016). Auch das heutige bekannte Modell der Ziffernnote für jedes Kind ist im Laufe der Zeit entstanden. Welche Eigenschaften oder Leistungen wie bewertet und gewichtet wurden, änderte sich im Laufe der Zeit.

Die Ziffernnote ist eine Form der Quantifizierung von Messen und vermittelt eine Intervallskalierung, in welcher es erlaubt ist, den Mittelwert (M) zu bilden. Ein solches Skalenniveau setzt jedoch voraus, dass alle Abstände zwischen den Noten gleich groß sind und somit der Abstand zwischen der Note 1 und 2 gleich groß ist, wie der Abstand zwischen der Note 3 und 4. Noten werden jedoch multidimensional aus mehreren Leistungen, Beobachtungen und inneren Eindrücken der Lehrkraft gewonnen und bestehen neben Leistungen auch aus dem Sozial- und Arbeitsverhalten. So beeinflusst das soziale Verhalten beispielsweise in einer Studie in deutschen Klassen der Sekundarstufe mehr die Zusammensetzung der Noten als das gemessene Kompetenzniveau des Unterrichtsfaches (De Vries et al., 2018).
Ziffernnoten sind daher eher als Ordinalskala zu betrachten, da die Noten in Rängen aufsteigen, jedoch die Abstände zwischen den Noten nicht gleich sind. Dies bedeutet, dass aus den Noten kein Mittelwert gebildet wird, sondern die Verteilung in Prozenträngen (10, 25, 50, 75, 90) darstellen wird. Der Median ist dabei der Prozentrang 50. Auch wenn Lehrkräfte Proben konstruieren und dabei Verrechnungsschlüssel als sogenannte Notenschlüssel (60% = Note XY) verwenden, werden diese Verrechnungen nicht empirisch geprüft. Dadurch sind die Abstände zwischen den Noten nicht gleich groß und durch Messfehler verzerrt. Meist ist der Abstand zwischen den Noten 4 und 5 sehr viel größer als der Abstand zwischen der Note 1 und 2. Die Ziffernnote ist für die pädagogische Diagnostik als ungefähre Einschätzung des Leistungsstandes und des Arbeitsverhaltens der vergebenden Lehrkraft zu interpretieren, wobei nachzufragen ist, auf welcher Basis und welchen Kriterien die Note besteht (siehe Beurteilungsfehler).

Arbeitsauftrag:
Erklären Sie in eigenen Worten, wann Noten den Anspruch an eine Intervallskalierung erfüllen würden.

Take Home Message:
In der pädagogischen Diagnostik werden intervallskalierte Skalen benötigt, um Unterschiede zwischen Personen vergleichen und interpretieren zu können. Noten erfüllen die Anforderungen an eine Intervallskalierung nicht.

Ziffernnoten sind situativ beeinflusst, werden meist selten nach Kriterien vergeben und bilden häufig den sozialen leistungsbezogenen Vergleich zwischen den Schüler:innen einer Klasse (soziale Bezugsnorm) ab. Daher wird die Vergabe von Ziffernnoten von Wissenschaftler:innen und mittlerweile auch von einigen Praktiker:innen generell kritisch gesehen und Wortgutachten anhand nachvollziehbarer kriterialer Bewertungsraster gefordert. Insbesondere in den unteren Jahrgangstufen (insbesondere in den Jahrgangstufen 1 und 2) wurden daher Lernentwicklungsgespräche anstelle von Zeugnissen mit Ziffernnoten eingeführt, mit dem Ziel, dass das Kind und die Eltern die Einschätzung des Leistungsstandes und der Entwicklung nachvollziehen können und mit der Lehrkraft gemeinsam zukünftige Ziele festlegen können. Ebenso wie bei der Ziffernnote besteht auch bei einem Wortgutachten die Frage, worauf die Bewertungen begründet sind.
Lehrkräfte gestalten Tests selten nach Erfassung eines latenten Merkmals einer Kompetenz und Teilkompetenz. Häufig werden ähnliche Aufgaben mit mittlerer Schwierigkeit aus

verschiedenen Inhaltsbereichen als Schulaufgabe verwendet. Hinzu kommen verschiedene Instruktionen, welche meist genau erfüllt werden müssen. Dies kann den Tests dann soweit verzerren, dass obwohl die Aufgaben eine Kompetenz messen würden, eher die Sorgfalt im Arbeitsverhalten oder auch das Verständnis der Aufgabentypen der Schüler:innen erfasst wird. Solche Tests sind nicht valide und vermischen meist unbeabsichtigt mehrere Kompetenzbereiche (siehe Kapitel 4.6).

Take Home Message:
Klassenlehrkräfte bewerten Schüler:innen meist im Kontext des Unterrichts als Teil einer Klasse. Alltag und Leistungsvergleiche in der Klasse gehen vielfach implizit in diese Bewertung mit ein. Dieser ganzheitliche Blick ist für die Diagnostik kritisch, da der/die Schüler:in individuell als einzelne Person in Hinblick auf eine diagnostische Fragestellung bewertet werden sollte.

Für die pädagogische Diagnostik gibt es somit **quantitative Informationen**, welche aus Zahlen bestehen. Diese Zahlen wurden anhand von Beobachtungen, Aufgaben, informellen oder standardisierten Tests gebildet. Es besteht die Frage, wie diese Zahlen gebildet wurden, welche Verrechnungsschritte benutzt wurden und was diese Zahlen repräsentieren. Ebenso gibt es **qualitative Informationen**, welche aus Gutachten, Zeugnissen, Einschätzungen, Bewertungen, Beobachtungen oder Gesprächen in Form des gesprochenen oder geschriebenen Wortes vorliegen. Die Kunst der pädagogischen Diagnostik ist es, diese Fülle an Informationen systematisch und nachvollziehbar für die Beantwortung einer pädagogischen Fragestellung heranzuziehen, zu gewichten und zu interpretieren, um auf dieser Grundlage eine **datenbasierte Entscheidung** (DBDM) zu treffen.

Videos:
Skalenniveaus (10:50)
Median, Mittelwerte, Boxplot und Normalverteilung (16:38)

Arbeitsauftrag:
Lesen Sie in weiterführenden Büchern nach, was Skalenniveaus sind und finden Sie Beispiele für Intervall-, Ordinal- und Nominalskalierungen.

Was sind quantitative und qualitative Merkmale sowie Auswertungen für die pädagogische Diagnostik?

4.1 Lernen und die Entwicklung von schulischen Kompetenzen

Menschen lernen und vergessen jeden Tag. Viele Lernprozesse passieren implizit und sind von der Person selbstgesteuert. In der Schule werden die explizit und von der Lehrkraft angeleiteten Inhalte geprüft und bewertet. Meist wird dabei nicht das Lernen an sich geprüft, sondern ein bestimmtes Produkt, eine mündliche Aussage oder eine schriftliche Leistung. Diese Bewertung erfolgt meist nach einer Kategorie oder einem Bewertungssystem, welches den Anforderungen des Lehrplans entsprechen soll. Diese Kategorien oder Bewertungssysteme von Lehrkräften sind jedoch nicht immer nachvollziehbar oder reliabel. Die Bildungsforschung versucht daher schulische Leistung und Lernen anhand von Instrumenten beispielsweise nach dem Literacy Konzept zu erfassen, um Aussagen über schulische Kompetenzen und deren Entwicklung zu treffen.

Allgemein angenommen wird, dass alle Kinder in der Schule lernen. Im Durchschnitt ist dies der Fall, für das einzelne Kind können die Kompetenzverläufe unterschiedlich sein. Theoretisch werden verschiedene Verläufe angenommen und auch empirisch in der Praxis gefunden:

Matthäuseffekt: „Denn wer da hat, dem wird gegeben, dass er die Fülle habe; wer aber nicht hat, dem wird auch das genommen, was er hat." (Mt. 25,29 NT). Je nach Vorwissen und personellen Fähigkeiten sowie Veranlagung lernen Kinder je nach Lernsituation unterschiedlich schnell. Würde die Schule somit jedes Kind bestmöglich nach seiner Begabung fördern, könnten in verschiedenen Lernbereichen ein Schereneffekt oder Matthäuseffekt entstehen. Insbesondere dann werden Schereneffekte angenommen, wenn Kinder mit Lernschwierigkeiten nicht unterstützt werden und Lehrkräfte sich insbesondere um leistungsstarke Kinder kümmern.

Kompensatorische Effekte der Schule: Auf der anderen Seite könnten kompensatorische Effekte begründet werden, in welchem insbesondere die leistungsschwächeren Schüler:innen gefördert werden. Diese holen dann im Vergleich zu ihren Mitschüler:innen auf (Hermann et al., 2022).

Effekt zur Mitte durch kompensatorische und bremsende Effekte: Die Schule vermittelt und prüft sehr verschiedene und spezifische Inhalte und differenziert je nach didaktischen Modellen unterschiedlich. So würde theoretisch ein Unterricht ohne Differenzierung nur ein Lernniveau vermitteln und prüfen. Kinder über oder unterhalb dieses Lernniveaus werden unter- oder überfordert und deren Entwicklung vermutlich unzureichend geprüft.

Stabile Entwicklung: Die Lernentwicklung ist für alle Schüler:innen gleich. Die Verteilung der Schüler:innen hinsichtlich der Lernentwicklung bleibt gleich. Leistungsstarke Schüler:innen bleiben leistungsstarke Schüler:innen und Kinder mit Lernschwierigkeiten verweilen am unteren Ende der Verteilung.

Anzunehmen ist, dass aufgrund der vertieften Inhalte des Lehrplans, der Reifung der Person sowie von Sozialisationseffekten eine durchschnittliche Kompetenzentwicklung auf Ebene aller Schüler:innen stattfindet (Siehe unterschiedliche Aufgaben bei Vera für Primarstufe und Sek I). Studien zur Kompetenzentwicklung sind Längsschnittstudien mit mehreren Schulleistungstests, die verschiedene Leistungsstufen der Klassenstufe und der Heterogenität der Schüler:innengruppe abdecken. So zeigte Hermann et al., 2022 für die Grundschule in Deutschland mit Daten des Nationalen Bildungspanel (NEPS) einen leichten kompensatorischen Effekt zwischen den Jahrgangstufen 1 und 4.

Komplexer wird die Frage nach der Lernentwicklung, wenn Schüler:innen mit sonderpädagogischen Unterstützungsbedarf miteinbezogen werden. Diese Schüler:innen verfügen nur bedingt über die gleichen Kompetenzen wie ihre Mitschüler:innen. Während SchülerInnen mit sonderpädagogischen Unterstützungsbedarf mit gleichem Lehrplan ähnliche Leistungen

erbringen, sind die Ergebnisse im Unterstützungsbedarf Lernen meist eine Standardabweichung geringer. So erreichen Schüler:innen im Unterstützungsbedarf Lernen in Förderschulen nach der 9. Klasse keine ausreichenden Basiskompetenzen und befinden sich auf den PISA Leveln 1 und 2 (Gebhardt et al., 2015b). Dies bedeutet, dass mathematische Aufgabenstellungen wie der Umgang mit Platzhaltern (Variablen), komplexere Sachaufgaben und der Umgang mit Division und Brüchen auch in der neunten Klasse eine Herausforderung sind (Gebhardt et al., 2013).

Take-Home-Message:
Die Messung der Entwicklung hängt von der gemessenen Fertigkeit oder Kompetenz ab. Angenommen wird, dass eine Fertigkeit (Skill) schneller gelernt wird als eine Kompetenz eines Faches. Schulisches Lernen wird in der Forschung unterschiedlich definiert und je nach Fragestellung anders gemessen (Siehe Kapitel 4).

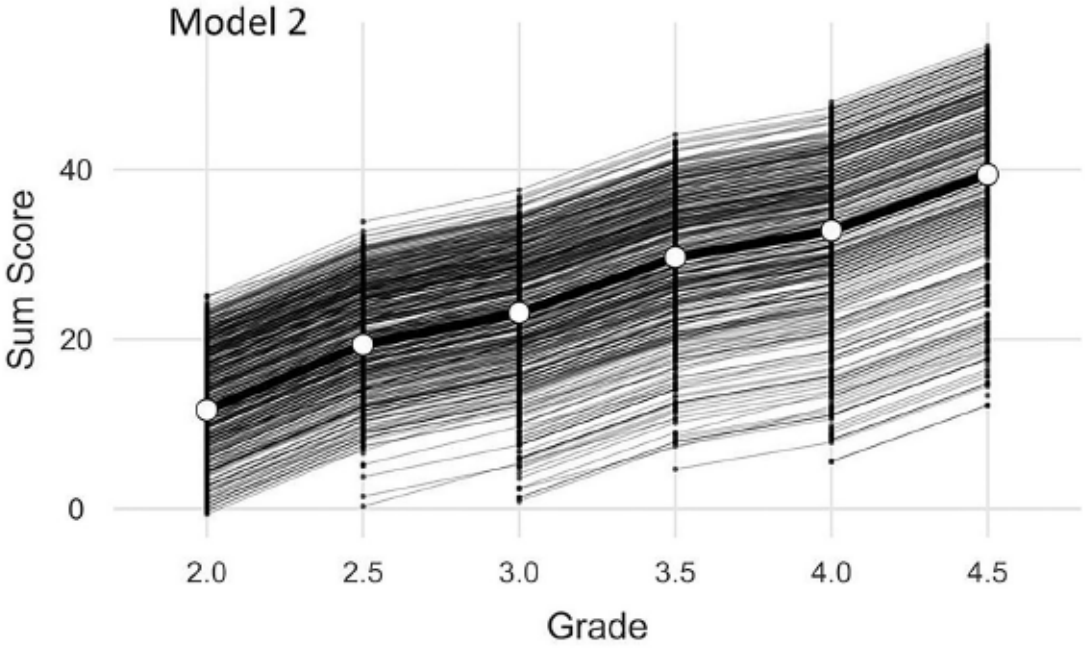
Für Vergleiche über die Zeit sind mehrere Messungen mit vergleichbaren Instrumenten notwendig. Hierzu gibt es nur wenige Studien, welche über verankerte Itempools verfügen, und vergleichbar sind. Im Rahmen der Studie KESS in Hamburg wurden 23 Förderschulen miterhoben und in der Arbeit von Schwanenberg (2009) verglichen. Die Tabelle 1 gibt ausgewählte Ergebnisse von Schwanenberg (2009, S. 81) wieder. Man erkennt, dass die 278 Förderschüler:innen niedrigere Mittelwerte als die 8983 Schüler:innen der Regelschule hatten, aber die Effektstärke Cohens d in einem ähnlichen Bereich liegt. Dies bedeutet, dass die Förderschüler:innen zwischen Klasse 7 und 9 eine ähnliche Verbesserung der Ergebnisse hatten wie die Schüler:innen der allgemeinen Schule zwischen Klasse 4 und 7. Da die verschiedenen Schülergruppen unterschiedliche Lehrpläne und in unterschiedlichen Klassenstufen waren, ist ein solcher Vergleich mit ähnlichen Tests schwierig zu interpretieren. Festzuhalten ist jedoch, dass die schulischen Leistungen sehr heterogen sind.

Tabelle 1
Mittelwerte von KESS allgemein und in der Förderschulerhebung KESSF

		KESS 4 (Klasse 4)	KESS 7 (Klasse 7)		
Schulart	*N*	*Mean (SD)*	*Mean (SD)*	*Diff*	*d*
Alle SchülerInnen	8983	449 (98)	506 (100)	57	.57
		Klasse 7F	Klasse 9F		
Förderschule Lernen	278	320 (71)	363 (70)	43	.61

Ein ähnliches Bild zeigt die Auswertung von Schurig et al. (2022) von der Rechtschreibentwicklung von der 2. bis 4. Klassenstufe im Längsschnitt. Halbjährlich wurde der Rechtschreibtest Rainer durchgeführt. In Abbildung 1 ist zu erkennen, dass die Entwicklung der Rechtschreibkompetenz stabil ist und sich alle Kinder ähnlich zwischen den Testzeitpunkten verbessern. Somit blieben auch die letzten 10% der Schülerschaft stabil und konnten nicht aufholen. Der Test Rainer hat sehr viele einfache Aufgaben und misst daher auch in den Klassenstufen 3 bis 4 im unteren Bereich relativ differenziert.

Abbildung 1
Entwicklung der Summenscores im Test Rainer der Klassenstufen 2 bis 4 von Schurig et al. (2022)



Die Frage der Kompetenzverläufe hängt mit den verwendeten Kompetenzmodellen und deren Instrumenten zusammen. Aktuell kann man festhalten, dass unser Bildungssystem keine starken kompensatorischen Effekte aufweist und aktuell nicht alle die Schule mit ausreichenden Kompetenzen für den beruflichen Bereich verlassen. Eine Verbesserung des Systems durch direkte Unterstützungsmaßnahmen, mehr Individualisierung und eine bessere Verbindung von Diagnostik und Förderung sind notwendig (siehe Kapitel 7).

Take-Home-Message:
In der Öffentlichkeit und Politik herrscht die Annahme, dass alle Kinder innerhalb eines Schuljahres Lernzuwächse haben. In Bezug auf die Kompetenzentwicklung stimmt dies nur bedingt, da im aktuellen Schulsystem insbesondere Kinder mit Risikoprofil nicht immer von Unterricht und Förderung profitieren. Um die Entwicklung von allen Kindern sicher zu stellen, wird pädagogische Diagnostik und damit verbundene pädagogische Unterstützung möglichst auch auf individueller Ebene in allen Schulen benötigt.

Weitere Informationen zum Bereich Kompetenzen:
Nationaler Berichtsband von PISA 2018
IQB Bildungstrend 2021

Exkurs:

Die Frage, ob die Bewertungen von Unterricht auch mit einem wirksamen Unterricht (mit nachweisbarem Lernerfolg) einhergehen, versucht der Podcast Psychologie fürs Klassenzimmer von Dr. Benedikt Wisniewski zum Thema Unterrichtsbeobachtungen zu beantworten.

4.2 Beobachtung zum Lernen und qualitative Fehleranalyse

Kinder und Jugendliche lernen im Unterricht und meistern mehr oder weniger erfolgreich verschiedene Aufgaben. Die Kompetenz der Lehrkraft liegt darin, die passenden Aufgaben zu wählen und anhand der Ergebnisse und des beobachteten Lernprozesses Rückschlüsse zu ziehen. Aufgaben und Anforderungen im Unterricht können anhand des Lehrplanes und nach fachdidaktischen, psychologischen und pädagogischen Modellen als gestuft betrachtet werden. Verfügt eine Lehrkraft über eine hohe diagnostische Kompetenz und didaktisches Fachwissen, kann die Lehrkraft die Kompetenz des/der Schüler:in anhand von Aufgaben einschätzen.
Dies gelingt in der Praxis jedoch eher selten. Eine grobe Einteilung der Schüler:innen in Leistungsgruppen ist möglich, aber eine genaue Überprüfung des Leistungsstandes zum aktuellen Unterricht für die individuelle Förderung findet nur in Ausnahmen statt. Eine individuelle Prüfung ist aufwendig, erfordert hohe diagnostische sowie fachdidaktische Kompetenzen und ist aufgrund der fehlenden Standardisierung fehleranfällig.

Die individuellen Lösungswege des Kindes werden analysiert und mit Bezug zu einem Entwicklungsmodell oder Kompetenzstufenmodell bewertet. Das Modell ist notwendig, um zu bestimmen, in welcher Stufe die aktuelle Entwicklung des Kindes ist und welche Stufe als nächste erfolgt. Diese Modelle sind somit die gleichen oder ähnliche Modelle, welche hinter der Entwicklung von standardisierten Instrumenten liegen. Im Unterschied zu standardisierten Instrumenten, bei denen meist eine Aufgabe als richtig oder falsch ausgewertet werden, werden qualitative Instrumente nach der Art der Lösung betrachtet und der Lösungsweg analysiert. Die Frage ist, welche **Strategie** der/die Schüler:in verwendet hat. Falls die Lösung falsch ist, ist zu fragen, wie kam der/die Schüler:in zur falschen Antwort. Welche Denk- und Handlungsmuster hat der/die Schüler:in.

Eine Möglichkeit ist hier den/die Schüler:in zu bitten, seine/ihre Gedanken beim Lösungsweg laut zu formulieren (Think-aloud).

Weitere Informationen:
Beobachtungsbögen vom hessischen Bildungsserver
Beobachtungsbögen nach dem Konzept PIKAS

4.3 Diktate und Proben als Prüfungsform

Die Prüfungsform der **Diktate** wird in der Schule immer noch gerne eingesetzt. Die Frage ist jedoch, was diese Prüfungsform misst und ob dieses Ergebnis replizierbar ist. Eine Lehrkraft diktiert Wörter oder Text. Schüler:innen hören das gesprochene Wort und schreiben es auf. Die Lehrkräfte korrigieren die Rechtschreibung, zählen meist die Fehler und geben anhand eines vorher überlegten Schemas eine Rückmeldung meist verbunden mit einer Note.

Arbeitsauftrag:
Welche Formen des Diktates kennen Sie aus Ihrer Schulzeit in der Primarstufe oder Sekundarstufe? Wie häufig waren diese Diktate? Welche Rückmeldung haben Sie bekommen? Wann könnte ein Diktat objektiv und fair messen?

Entwerfen Sie ein möglichst faires Diktat mit fünf Sätzen für die 2. Klasse.

In der Forschung werden Diktate aus verschiedenen Gründen eher kritisch gesehen, da meist das Diktat nicht viel mit dem aktuellen Rechtschreibunterricht zu tun hat und die Prüfungsform weder objektive noch reliable Ansprüche einer fairen Messung erfüllt. Meist ist es auch für Kinder mit vielen Fehlern schwer bis unmöglich, sich zu verbessern und somit führt ein standardisiertes Diktat nur zu einer ungenauen Messung der sozialen Bezugsnorm.

Birkel (2009) wählte drei Diktate mit den Noten gut, befriedigend und ungenügend aus der Praxis aus und ließ diese von 415 Lehrkräften bewerten. Hierbei zeigten sich stark unterschiedliche Bewertungen der Fehler und somit auch bei der Vergabe der Zensuren. Ebenso wurde sehr streng bewertet, so dass bereits ab 8% Fehler im Bezug zur Wörteranzahl des Diktates die Note ungenügend erreichte wird. Birkel (2009) folgerte daraus, dass die von Lehrkräften für objektiv gehaltene Leistungsbeurteilung des Diktates alles andere als genau die Rechtschreibleistung der Schüler:innen misst und insbesondere die Konstruktion ohne systemische Regeln und Auswertungen kritisch zu sehen ist.

Take Home Message:
Wird aus einer Probe, Klausur oder Test ein Summenwert gebildet und eine Note errechnet, dann gelten auch für selbstentwickelte Leistungserhebungen die Anforderungen der Testtheorie. Verletzt man deren Anforderungen, sollte man sich dessen bewusst sein und die Ergebnisse von selbstkonstruierten Proben nur eingeschränkt interpretieren.

Ein Diktat kann jedoch auch anhand eines Grundwortschatzes und anhand konkreter Rechtschreibschwierigkeiten entwickelt werden. Ziel eines solchen Tests ist es, die Rechtschreibung zu messen und Lehrkräften rückzumelden, in welchem Bereich der Rechtschreibung das Kind schon sicher ist oder noch weiterarbeiten muss. Daher wird häufig zwischen verschiedenen Bereichen wie der alphabetischen Rechtschreibstrategie („Ich kann so schreiben wie ich höre.", Mitsprechwörter), der morphematischen Rechtschreibstrategie („Ich kann durch Nachdenken und das Anwenden von Strategien das Wort richtig schreiben z. B. Hund (ich höre ein ‚t' am Ende) – Hun-de (ich höre in der zweiten Silbe ein ‚d').", Nachdenkwörter) und der orthographischen Rechtschreibstrategie („Ich muss mir einprägen, wie man das Wort schreibt.", Lernwörter) unterschieden. Die Tests werten dann meist nicht das Wort, sondern einzelne Graphemtreffer nach bestimmten Verrechnungsregeln aus. Die Tests werden dann in Studien anhand der Gütekriterien geprüft und gegebenenfalls verbessert. Auswertungen nach der individuellen, sozialen und kriterialen Norm sind dann anhand des Manuals und einer gegebenenfalls angegebenen Normierung möglich. Strathmann et al. (2010) beschreiben ein solches Vorgehen in ihrer Studie mit 128 Grundschulkindern, deren Rechtschreibleistung mittels 20 parallelisierter Diktate über 20 Wochen gemessen wurde.

Ein freies Verfahren zur Messung der Rechtschreibung mittels Wortdiktaten liegt beispielsweise mit dem Rainer 1-4 für die Klassenstufen 1 bis 4 vor.

Arbeitsauftrag:
Verbessern Sie Ihr eigenes Diktat. Überprüfen Sie die einzelnen Wörter Ihres Diktattextes durch einen Lesbarkeitsindex LIX beispielsweise mit dem Tool Ratte oder auf der Homepage von Lenhard & Lenhard.

Die Leseprobe

Frau Müller benötigt eine Leseprobe. Sie nimmt einen mittelschweren Lesetext der dritten Klasse und fügt verschiedene Übungen und Aufgaben zum Test hinzu. Dabei verwendet sie ein paar Standardübungen, welche der Klasse bekannt sind und neue Formate, welche die Klasse bisher kaum verwendet hat.

Bei der Bewertung stellt sie verwundert fest, dass sie viele Kinder mit relativ schlechten Noten bewertet. Denn ein Teil der Kinder mit hoher Lesekompetenz löst nur gering das ungeübte Aufgabenformat. Frau Müller hat neben der Schwierigkeit des Lesetextes auch die Bekanntheit der Aufgabenformate als Schwierigkeit im Test miteingebaut. Scheitern bereits einige Kinder an der Aufgabenstellung, dann kann der Test nicht fair messen (siehe Kapitel 4.6). Sinnvoll für Proben ist es daher, neben der Art des Textes auch das Aufgabenformat im Unterricht zu besprechen und die Strategien einzuüben, welche zur Lösung benötigt werden. Aus diesem Grund beinhalten auch psychometrische Tests Beispielitems, damit sichergestellt wird, dass das Aufgabenformat bekannt ist. Für Lehrkräfte ist es auch mit bekannten Strategien und Aufgabenformaten kein Problem leichte, mittlere und schwere Aufgaben (Items) zu entwickeln, um zwischen den Kindern differenzierende Ergebnisse zu bekommen. Um eine gute Note in der Probe von Frau Müller zu erhalten, muss man neben der Lesekompetenz auch über eine gute Schreibkompetenz verfügen. Somit misst die Probe von Frau Müller nicht nur eine Dimension, sondern mehrere Dimensionen. Die Ergebnisse der Probe von Frau Müller kann somit kaum interpretiert werden und es ist nicht sicher, was die Note der Probe genau beinhaltet.

Take Home Message:
Proben oder Klausuren, bei denen Kinder an Aufgabenformaten, versteckten Tricks oder sehr nah formulierten Antwortalternativen scheitern, sind unfair und messen eher Vorwissen und das Arbeitsverhalten als das gewünschte latente Konstrukt.

Exkurs:
Die Frage nach dem Sinn von Korrekturen versucht der Podcast Psychologie fürs Klassenzimmer von Dr. Benedikt Wisniewski zum Thema Korrekturen zu beantworten.

4.4 Latente Konstrukte und Dimensionen

In der Pädagogik sind einzelne Situationen oder Ereignisse beobachtbar, aber nicht die dahinterliegende Kompetenz. So kann die didaktische Situation und Lernsituation direkt beobachtet werden, aber nicht den einzelnen Lernprozess bei den betroffenen Personen. Der Lernprozess ist ein aktiver Prozess der betroffenen Person, welcher für Außenstehende im Verborgenen stattfindet und auf den nur indirekt anhand von Verhaltensspuren geschlossen werden kann. Die meisten wesentlichen Konstrukte in der Pädagogik kann man daher meist nur als **latente** Konstrukte anhand von mehreren Beobachtungen oder Aufgaben messen. Manifeste Konstrukte sind dagegen die einzelnen konkret vorliegenden Beobachtungen und Aufgaben. Die Messung des Körpergewichts ist beispielsweise als **manifest** zu betrachten, welches durch eine Messung mit der Waage abgeleitet werden kann. Die körperliche Fitness dagegen ist ein latentes Konstrukt, für welches mehrere verschiedene Messungen benötigt werden.

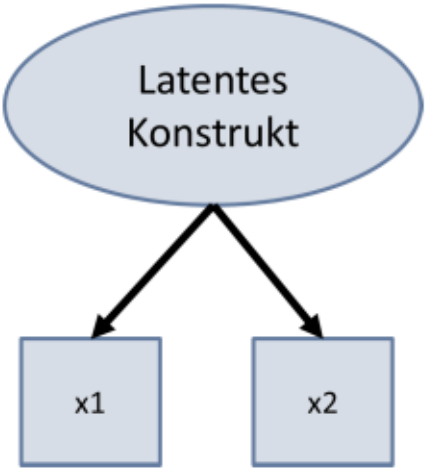
Arbeitsauftrag:
Sie unterrichten Schwimmen in der Grundschule. Manche Kinder waren noch nie im Schwimmbad und andere Kinder haben bereits ihr Seepferdchen nach Kriterien der DLRG. Welche Kompetenzen und Teilkompetenzen sollen die Kinder in ihrem 15-teiligen Schwimmkurs erwerben? Wie könnte man diese Teilkompetenzen beobachten? Begründen Sie nach welcher Bezugsnorm Sie die Kinder abschließend bewerten würden und warum.

Schulische Leistungen sind latente Konstrukte, welche durch Tests und Beobachtungen anhand mehrerer, unterschiedlich schwerer Aufgaben gemessen werden. Entspricht die Auswahl dieser Aufgaben einem kompetenzorientierten Vorgehen und bilden diese einen Teil eines Literacy Konzeptes in Form eines Kompetenzstufenmodells ab, kann man auch von einem Kompetenztest sprechen. Entspricht der Test den Anforderungen der psychometrischen Gütekriterien (siehe Kapitel 4.6), dann wird anhand von mehreren Aufgaben für die Person ein Wert gebildet, welcher zu einer hohen Wahrscheinlichkeit (meist 95%) dem wahrscheinlichen, sogenannten „wahren Wert" des latenten Merkmals der Person entspricht. Gute Tests geben hierbei ein Konfidenzintervall an, in dessen Bereich der wahre Wert mit der angenommenen Wahrscheinlichkeit (meist 95%) liegt. Diese Annahmen werden anhand der Messungenauigkeit und der angenommenen Fehlerwahrscheinlichkeit (meist 5%) begründet. Bei Messungen in wissenschaftlichen Studien werden daher meist ein empirischer Wert und der dazugehörige Standardfehler (SE) angegeben und es werden zur Verrechnung von latenten Variablen probabilistische Modelle verwendet.

Exkurs:
Ob ein Ergebnis zufällig ist oder nicht, ist die Frage der Interferenzstatistik. Umso mehr Aufgaben eine Person in einem Test bearbeitet, umso mehr Informationen stehen zur Verfügung und umso unwahrscheinlicher ist es, dass die Aufgaben nur zufällig gelöst wurden. R. A. Fischer (1935) stellte sich eine ähnliche Frage am Beispiel der teetrinkenden Frau. Diese Frau behauptet, dass es für sie einen Unterschied macht, ob zuerst Tee oder Milch in die Tasse geschüttet wird und sie treffsicher den Unterschied vorhersagen kann. Nun ist die Frage, wie würde ein passendes Experiment aufgebaut sein, um die Fähigkeit der Frau mit einer geringen Fehlerwahrscheinlichkeit zu bestimmen. Wie viele Durchgänge benötigt man, um mit Sicherheit den Zufall auszuschließen?

Wikipedia: Lady Tasting Tea
Video: Lady Tasting Tea (6:39)

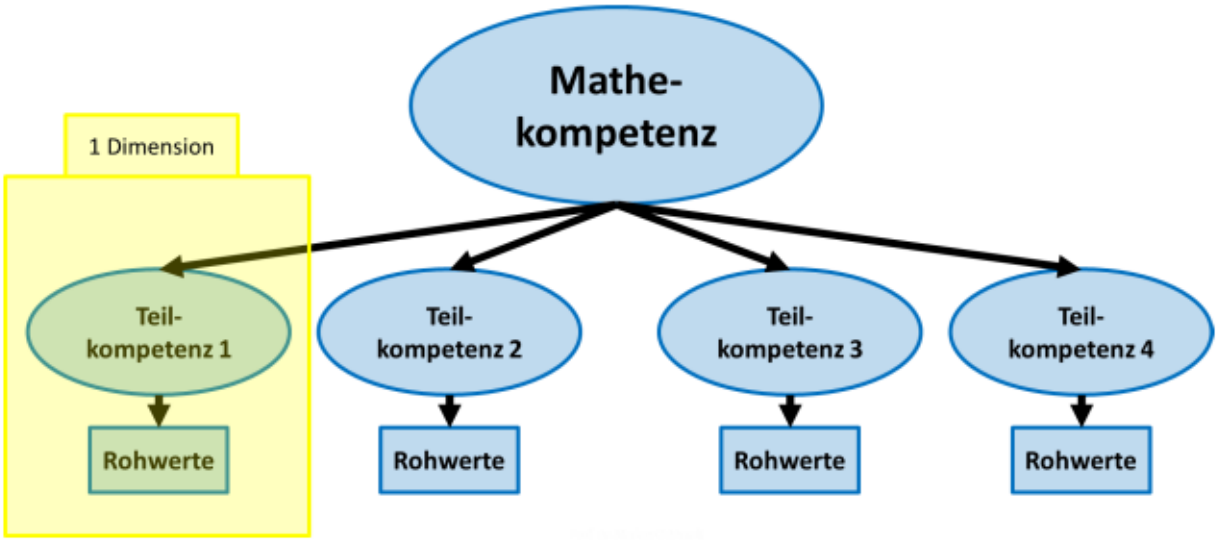
Abbildung 6
Strukturmodell eines latenten Konstrukts



Latente Variablen werden als Ellipsen in theoretischen **Strukturmodellen** dargestellt. Die manifesten Variablen werden als einzelne Rechtecke pro Item dargestellt sind. Die latente Fähigkeit wirkt auf das Antwortverhalten, daher ist die Pfeilrichtung in Abbildung 6 vom latenten Konstrukt auf die einzelnen Items gerichtet. Die Ebene der Items umfasst das empirische Messmodell, wobei jedes Item auch einen Messfehler beinhaltet. Tests werden somit nach einem theoretischen Modell entworfen und mittels eines empirischen Modells geprüft.

Schulleistungstests sind meist multidimensional und bestehen aus mehreren Subtests. Meist wird mit dem gesamten Test eine Kompetenz gemessen und die einzelnen Tests messen eine Teilkompetenz. Die Subtests haben eine Instruktion und bestehen aus ähnlichen Items. Pro Subtest werden dann die Rohwerte zusammengerechnet und ergeben dann das Profil einer Person. In Abbildung 7 wird ein Test in Mathematik dargestellt, welcher aus vier Dimensionen besteht. In Abbildung 7 ist anstelle von 20 Items pro Dimension vereinfacht nur ein Rechteck aufgeführt. Der Gesamtwert der Person ergibt sich aus den vier Subtests. Je nach angenommenem Testmodell können die Aufgaben und Dimensionen zur Verrechnung verschieden gewichtet werden. Da dies insbesondere bei Paper-Pencil-Tests jedoch herausfordernd ist, gehen meist alle Aufgaben zu 100% in die Verrechnung mit ein und es wird somit das Raschmodell (Rasch, 1961) angenommen.

Abbildung 7
Strukturmodell eines Tests für Mathematik



Der Test Demat 2+ besteht beispielsweise aus den zehn Subtests Zahleneigenschaften, Längenvergleich, Addition und Subtraktion, Verdoppeln und Halbieren, Division, Rechnen mit Geld, Sachaufgaben und Geometrie. Der Test misst die Leistung in Mathematik anhand der Lehrplaninhalte für das Fach Mathematik in der 2. Klasse. Dieser Test kann als Gruppentest für die 2. und 3. Klasse in 45 Minuten durchgeführt werden. Anhand der zehn

Dimensionen wird ein breites Profil der Schüler:innen gemessen und einzelne Stärken oder Schwierigkeiten werden dokumentiert. Dadurch sind Rückmeldungen für eine spezifische Förderplanung möglich.

Take Home Message:
Ein latentes Konstrukt wird mit meist mehreren Aufgaben (Items) gemessen. Die einzelnen Aufgaben werden als manifest betrachtet.

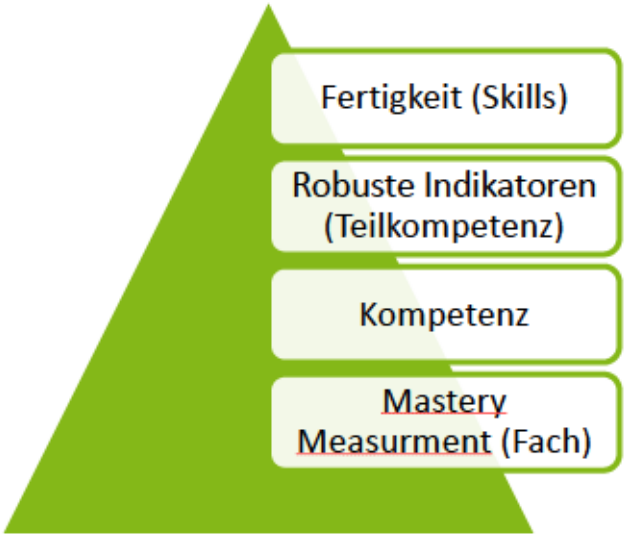
Umso mehr Dimensionen ein Test hat, um so breiter misst der Test als multidimensionaler Test auch das latente Merkmal. Ein Nachteil eines multidimensionalen Tests ist jedoch, dass jeder Subtest reliabel gemessen werden muss, um eine Aussage darüber zu ermöglichen. Daher wird für solche Tests auch mehr Zeit (meist mindestens fünf Minuten pro Dimension) benötigt. Aus diesem Grund verwendet man diese Tests bei einer umfänglichen Statuserhebung für ein Gutachten oder wenn man noch wenig Informationen über das Profil des/der Schüler:in hat. Bei wiederkehrenden Testungen lohnt sich dagegen ein so umfänglicher multidimensionaler Test nicht. Hier ist es sinnvoller, nur die für die Förderung relevanten Subtests oder Merkmale zu erheben. Meist korrelieren die Subtests in Studien recht hoch, so dass, falls kein Anfangsverdacht oder gegenteilige Informationen vorliegen, angenommen werden kann, dass die einzelnen Subtests im Regelfall ähnlich beantwortet werden. Hat eine Person jedoch ein sehr unterschiedliches Profil in der Beantwortung der einzelnen Subtests, ist dies ein wichtiger Ansatzpunkt für eine weitere genauere Beobachtung.

Videos:
Manifeste und Latente Konstrukte (13:47)
Unabhängige und abhängige Variablen (9:53)

Arbeitsauftrag:
Was ist ein manifestes und ein latentes Konstrukt? Erklären Sie dies in eigenen Beispielen. Leihen Sie sich einen mutidimensionalen Test aus der Testothek aus oder lesen Sie die Beschreibung des ELFE II.
Was sind die verschiedenen Dimensionen und ihre Zusammenhänge nach dem Testhandbuch? Was ist der Unterschied zwischen Rohwerten und normierten Werten?

Als weitere Möglichkeit kann man eine Dimension eines Tests selbst sehr unterschiedlich gestalten. Je nachdem welche Fertigkeit oder Kompetenz gemessen werden möchte, werden die Items nach sehr spezifischen Regeln gebildet, um eine Fertigkeit (Skill) möglichst genau zu erfassen oder um alle Kompetenzen eines Schuljahres zu einem einzelnen Schulfach zu messen. Je nach Ziel wird der Test somit unterschiedlich konstruiert, um den angestrebten Zielbereich abzudecken. In Abbildung 9 ist dieser Aspekt schematisch dargestellt. Wenn die Lehrkraft eine spezifische Entwicklung beispielsweise im Addieren bis 100 mit Zehnerübergang messen möchte, handelt es sich um einen Skilltest. Umfasst der Test mehr als eine einzelne Fertigkeit und wird auch über eine längere Zeit angewendet, kann man dies als Teilkompetenz auffassen. Von einem robusten Indikator wird gesprochen, wenn der gemessene Bereich des Tests für die Kompetenzentwicklung insgesamt sehr bedeutsam ist. So wird die Flüssigkeit des Kopfrechnens beispielsweise als ein bedeutsamer Indikator für Mathematik gesehen. Tests zur Messung der Kompetenz bestehen aus unterschiedlichen Aufgaben oder mehreren Dimensionen, um beispielsweise die mathematische Kompetenz in der 2. Klassenstufe zu messen. Eine noch breitere Anwendungsmöglichkeit ist es mehrere Kompetenzen zu messen, welche innerhalb eines Schuljahres in einem Fach verlangt wird. Hierzu können mehrere unterschiedliche Tests zu einem Assessment als Mastery Measurment zusammengenommen werden.

Abbildung 9
Messformen eines Tests



4.5 Latente Dimensionen im Fragebogen

Der Fragebogen DAKI ist ein Einschätzungsbogen zur diagnostischen Kompetenz für Lehrkräfte in der Inklusion. Lehrkräfte sollen anhand des Fragebogens einschätzen, in welchen Dimensionen sie sich qualifiziert fühlen und in welchen Dimensionen sie noch weitere Fortbildungen benötigen. Die Ergebnisse bei Sonderpädagog:innen zeigen (Jungjohann & Gebhardt, im Druck), dass insbesondere in der Dimension Lernverlaufsdiagnostik (im Englischen Progress Monitoring, PoM) noch Fortbildungsbedarf besteht. Bevor die Autor:innen jedoch die Summenwerte der einzelnen Dimensionen interpretieren durften, wurden diese im Artikel mittels korrelativer Ansätze geprüft. Die Annahme dabei ist, dass die theoretisch betrachteten ähnlichen Items auch empirisch höher miteinander korrelieren. Daher werden alle Zusammenhänge der Items mittels der Korrelation nach Pearson r betrachtet.

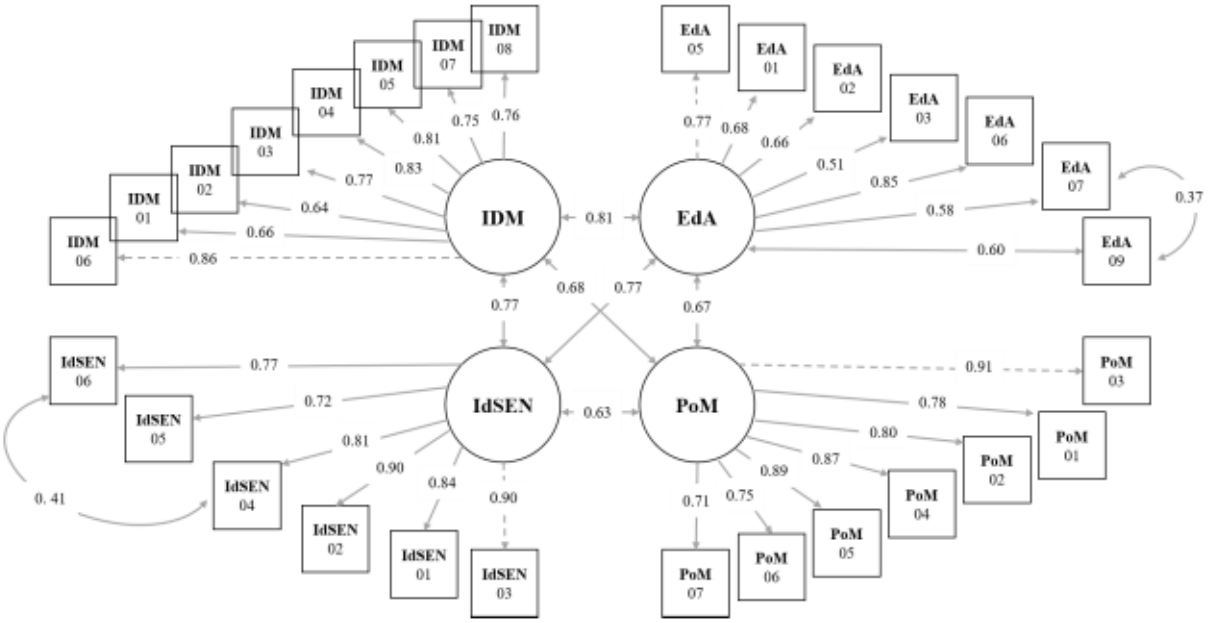
Um zu ermitteln, ob die Items zu einer Dimension zusammenpassen, wird häufig für eine Dimension der Cronbachs Alpha als Maß der internen Konsistenz angegeben. Hierbei wird die mittlere Korrelation der Items in Bezug auf ihre Anzahl betrachtet. Der Cronbachs Alpha sollte bei Fragebögen über $\alpha = .60$ und bei Leistungstests über $\alpha = .70$ oder höher liegen. Der Fragebogen DAKI hat eine interne Konsistenz von $\alpha = .92, .82, .92, .93$ in seinen vier Dimension. Zur Prüfung von mehrdimensionalen Fragebögen ist der Cronbachs Alpha jedoch nur der erste Schritt, da er die Korrelationen aller Items prüft und nicht zwischen Dimensionen im Einzelnen unterscheidet. Für die Prüfung eines vierdimensionalen Instruments, bei dem die Items nur eine Dimension erfassen sollen, wird eine Konfirmatorische Faktorenanalyse (CFA) benötigt, in welcher ein festgelegtes Strukturmodell empirisch geprüft wird.

In Abbildung 8 ist das empirische Messmodell des DAKI dargestellt. Als Ellipsen sind die vier Dimensionen dargestellt, welche untereinander zwischen .63 und .81 korrelieren. Hierbei erkennt man, dass einzelne Items nicht zu 100% von dem latenten Konstrukt erklärt werden, sondern nur zu einem gewissen Teil laden. Die latenten Konstrukte laden nur auf die jeweiligen einzelnen Items zwischen .51 und .91.

Take Home Message:
Die Ergebnisse zu latenten Konstrukten aus Fragebögen und Tests werden häufig anhand von Strukturgleichungsmodellen berechnet und dargestellt. Korrelationen sind hierbei

bidirektionale Pfeile und Ladungen monodirektionale Pfeile, welche bei standardisierter Darstellung zwischen 0 und 1 liegen.

Abbildung 8
Strukturmodell des DAKI



Video:
[Korrelationen (10:21)]

Arbeitsauftrag:
Warum laden die latenten Konstrukte unterschiedlich auf die einzelnen Items des [Fragebogens DAKI?] Was misst der Fragebogen und wie würden Sie sich selbst einschätzen?

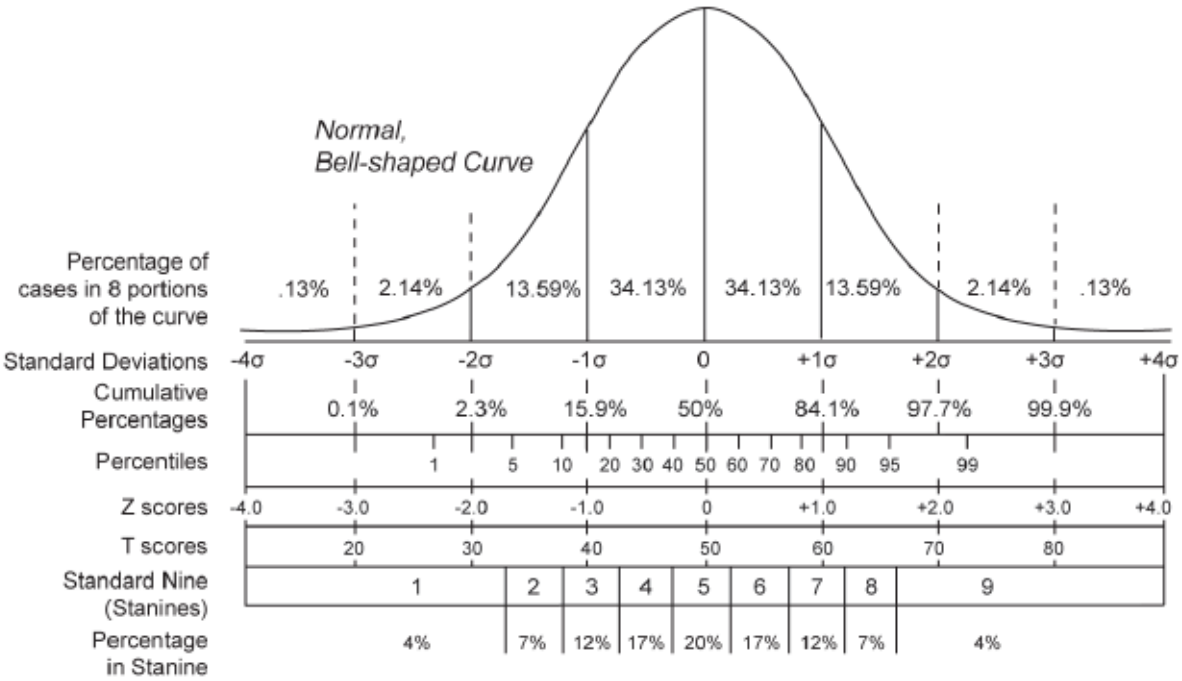
Mehr Informationen unter:
[Jungjohann, J., & Gebhardt, M. (2023). Dimensions of Classroom-Based Assessments in Inclusive Education: A Teachers' Questionnaire for Instructional Decision-Making, Educational Assessments, Identification of Special Educational Needs, and Progress Monitoring. International Journal of Special Education (IJSE), 38(1), 131–144.]

4.6 Normen

Standardisierte Schulleistungstests und Fragebögen sollen meistens einen Vergleich zur Klassenstufe oder Altersgruppe ermöglichen. Die individuell gelösten Aufgaben eines Tests ermöglichen die Auswertung eines Profils. Jedoch ist nur bei wenigen kriteriumsorientierten Tests festgesetzt, welcher Rohwert erreicht werden muss, um das Kriterium zu erreichen. Die meisten Schulleistungstests messen anhand einer sozialen Bezugsnorm. Sie ermöglichen einen direkten Vergleich, um festzustellen, dass die Mathematikkompetenz eines Kindes im Vergleich zu allen Kindern in Deutschland unter- oder überdurchschnittlich ist. Um diesen Vergleich zu ermöglichen, müssen Tests anhand größerer Studien normiert werden. Hierbei ist wichtig, dass in der Studie der gewünschte Vergleich, wie Alter oder Klassenstufe, ausreichend und repräsentativ vertreten ist. Das bedeutet für einen Test, welcher in allen Bundesländern einen Vergleich für die zweite und dritte Jahrgangstufe ermöglichen möchte, dass er auch theoretisch in alle Bundesländer in den zweiten und dritten Jahrgangstufen Ergebnisse benötigt und diese anhand der Gütekriterien prüft.

Für die Auswertung von Tests müssen die Anzahl der gelösten Aufgaben als Rohwerte in normierte Werte übertragen werden. Die Rohwerte sind die unstandardisierten Werte, welche je nach Aufgabenumfang des Subtests verschieden ausfallen können. Zur Interpretation des Tests muss entschieden werden, ob als Vergleich die Altersgruppe oder die Klassenstufe wählt wird. Anhand der in Normierungsstudien erstellten Normtabellen kann man dann den Rohwert der Person in einen Normwert transformieren. Dieser transformierte Wert stellt das Ergebnis der Person im Bezug zur gewählten Norm, also den Ergebnissen der Vergleichsstudie, dar. Ein Prozentrang 50, ein T-Wert 50 oder ein IQ 100 bedeutet, dass dieser Rohwert im Durchschnitt der Vergleichsgruppe liegt. Es gibt verschiedene Normskalen, wie die IQ-Skala oder die T-Wert Skala. Nachdem die gemessenen Konstrukte in der Normierungsstudie als Normalverteilung vorliegen, können anhand des Mittelwertes und der Standardabweichung der gesamten Studie durch verschiedene Berechnungsweisen Normen und Normtabellen erstellt werden. Diese standardisierten Werte sind, wie in Abbildung 9 dargestellt, vergleichbar. So liegt ein überdurchschnittlicher Wert über einer Standardabweichung vom Mittelwert, was einem IQ-Wert von 115 und einem T-Wert von 60 entspricht. Ein unterdurchschnittlicher Wert liegt unter einer Standardabweichung vom Mittelwert, was einem IQ-Wert von 85 oder einem T-Wert von 40 entspricht. Die normierten Werte können Sie auch mit Hilfe von Online-Rechnern umrechnen, wie beispielsweise dem Normwert-Rechner.

Abbildung 9
Normen aus Wikimedia



Take Home Message:
Für die Interpretation von Werten ist immer der Abstand zur Mitte (Prozentrang 50) der Verteilung entscheidend. Entweder entspricht die Leistung des Kindes den durchschnittlichen Leistungen aller Kinder oder er weicht über-/unterdurchschnittlich ab.

In Tabelle 5 wurden für den Demat 2+ die Rohwerte in Normwerte und Prozentränge übertragen. Diese Werte darf man aber nicht als einzelne Werte, sondern eher als Bereich interpretieren. Denn jede Messung von latenten Konstrukten basiert auf Annahmen zur Wahrscheinlichkeit und beinhaltet somit auch den Messfehler. Daher wird neben den normierten Werten auch ein Standardfehler (SE) angegeben, welcher vom normierten Wert abgezogen oder addiert wird, um das Konfidenzintervall zu erhalten. In diesem Bereich wird zu einer festgesetzten Wahrscheinlichkeit von meist 95% der wahre Wert der Person

vermutet. Die jeweiligen Subtests des Demat 2+ bestehen jedoch nur aus wenigen Items, so dass eine reliable Messung nur für den Gesamtwert vorliegt. Der Gesamtwert selbst steht für die Leistungen in Mathematik anhand des Lernplans. Inwieweit und ob man diesen Wert auch als Wert für die mathematische Kompetenz der Kinder betrachten kann, ist eine theoretische Diskussion. Je nach gewähltem Modell können andere Subtests oder Gewichtungen beim Verrechnen der Werte erfolgen. Gesamtwerte sind somit immer auch die Operationalisierung eines dahinterliegenden theoretischen Modells, welches für die Interpretation des Wertes beachtet werden muss.

Tabelle 5
Demat 2+ eines fiktiven Kindes Angelo in der 2. Klasse

Subtest	Rohwerte*	Prozentrang		
Zahleneigenschaften	1 von 2	55		
Längenvergleich	2 von 4	32		
Addition	2 von 4	50		
Subtraktion	2 von 4	44		
Verdoppeln	2 von 3	33		
Halbieren	2 von 3	31		
Division	3 von 4	52		
Rechnen mit Geld	2 von 4	43		
Sachaufgaben	3 von 4	60		
Geometrie	2 von 4	60		
			T-Wert	T-Wert Bereich
Gesamt	21 von 36	48-53	50	47-53

*Rohwert werden als Anzahl der gelösten Aufgaben angegeben

Weitere Informationen:
Kapitel 4 insbesondere ab Seite 99 bei Bundschuh und Winkler (2019).

Video:
Normen bei Tests 17:48

Arbeitsauftrag:
Erklären Sie den Begriff Standardabweichung in eigenen Worten.
Was ist der Unterschied zwischen Rohwerten und normierten Werten?
Welchem T-Wert entspricht ein z-Wert von -3?
Welchem T-Wert entspricht ein IQ-Wert von 160?
Welchem T-Wert entspricht ein IQ-Wert von 155?
Take Home Message:

Normen sind immer ein sozialer Vergleich zu einer Altersgruppe oder Schulklasse. Normen repräsentieren daher keine kriteriale Bezugsnorm. Ob oder inwiefern der soziale Vergleich angemessen und notwendig ist, ist eine pädagogische Fragestellung. Selbst wenn eine Person nur über Fähigkeiten verfügt, welche nur 2% der Bevölkerung haben, ist zu fragen, ob es deswegen ein pädagogisches Problem gibt und inwiefern dieses Problem nicht durch die Umwelt oder weitere Faktoren gelöst werden kann. Ein Ergebnis außerhalb der Norm muss daher immer in einen Gesamtzusammenhang betrachtet werden und reicht als alleiniges Merkmal nicht aus für eine pädagogische Entscheidung.

Übungsaufgabe Normen:

Pro Zeile werden die Normwerte verrechnet. Ergänzen Sie die fehlenden Werte pro Zeile.

Prozentrang	T-Wert	IQ-Wert	z-Wert	SW-Wert
50	50	____	0	100
____	40	85	_____	90
____	____	77,5	-1,5	85
97.72	70	____	____	____
99,87	____	145	____	130

Zu Prüfung Ihrer Lösung wird der Normwertrechner von psychometrica.de empfohlen

4.7 Grundlagen der Testtheorie

Die klassische Testtheorie (KTT; Lord & Novick, 1968) ist die Grundlage der Gütekriterien. Sie besteht aus mehreren Axiomen. Ein Axiom ist eine in einer Theorie als gültig und wahr angenommene Aussage, die keines Beweises bedarf. Im ersten Axiom wird festgelegt, dass eine beobachtete Merkmalsausprägung X einer Person sich aus den Komponenten der wahren Merkmalsausprägung T (true score) und einem zufälligen Messfehler E zusammensetzt.

$$E + T = X$$

Die KTT geht davon aus, dass die wahre Merkmalsausprägung intraindividuell konstant ist und die Messfehler sich bei wiederholten Messungen herausmitteln. Somit wird erwartet, dass bei sehr vielen Messungen der Messfehler irgendwann dem Wert 0 entspricht und die beobachtete Merkmalausprägung X dem wahren Wert der Person E entspricht. Das ist ein Grund dafür, warum zur Erfassung der Ausprägung eines Merkmals bei einer Person mehrere Items verwendet werden, die zusammen ein Konstrukt erfassen. Jedes dieser Items wird dabei als wiederholte Messung aufgefasst. Messfehler werden durch die Zusammenfassung der einzelnen Messungen zu einem Testwert ausgeglichen. Nach der KTT ist der Messfehler E unabhängig von der wahren Merkmalsausprägung T. Ebenso sind die Messfehler E bei jeder Messung unabhängig voneinander. Aus den Annahmen der KTT wurden die Reliabilität und die Gütekriterien abgeleitet (Moosbrugger, 2012). Im oberen Beispiel des DAKI in Abbildung 8 wird somit die Anforderung der KTT durch die Korrelationen jeweils zweier Items im Strukturmodell verletzt.

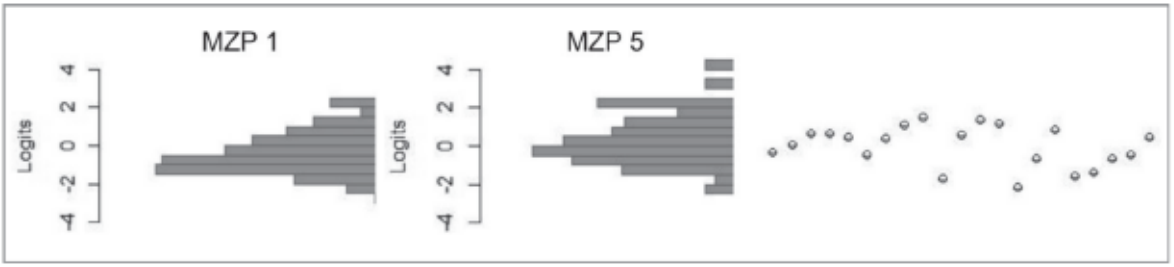
Take Home Message:
Jede Messung beinhaltet auch einen Messfehler. Bei Auswertungen von psychometrischen Instrumenten muss immer der Messfehler, meist in Form des Konfidenzintervalls, beachtet werden.

Diese strengen Annahmen lassen sich jedoch mit den statistischen Verfahren der KTT nur eingeschränkt prüfen. Insbesondere im Bereich der Leistungs- und Kompetenztests wird die KTT mit der probabilistische Testtheorie (Item-Response-Theorie (IRT)) erweitert. Bei den Item-Response-Modellen wird **nicht** das Testverhalten mit dem gemessenen Verhalten gleichgesetzt, sondern die beobachtete Variable nur als Indikator gesehen, welches ein Auftreten erhöht. Daher spricht man auch von *latent-trait-theories* und *probabilistischen Modellen* (Rost & Spada, 1978). Zentral für die Feststellung der Reliabilität von IRT Modellen ist die Feststellung der Präzision, mit der ein Testwert anhand einzelner Lösungswahrscheinlichkeiten ermittelt wird. Diese differiert dabei je nach der Lage im Leistungskontinuum. Die Genauigkeit der Messung ist im mittleren Bereich besser als im extremen Bereich (Rost, 2004). Das Raschmodell und seine Erweiterungen werden daher mittlerweile häufig bei der psychometrischen Prüfung und Skalierung von Leistungstests und Kompetenztests verwendet. Insbesondere bei Fragen zur Fairness von Tests für bestimmte Gruppen oder auch bei Tests zur Messung des Lernverlaufs werden probabilistische Modelle benötigt.

Bei IRT Modellen werden die Schwierigkeiten der Items (Itemparameter) zuerst und in einem zweiten Schritt die Werte der Personen (Personenparameter) geschätzt. Dadurch kann man die Passung des Tests zu den einzelnen Personen prüfen. Häufig dargestellt wird das in einer Personen-Item-Map. In Abbildung 10 sind die Itemparameter von Gebhardt et al. (2015a) rechts als Punkte dargestellt. Der Itemparameter ist die Schwierigkeit, mit welcher 50% der Personen das Item lösen können. Die Punkte sind die einzelnen Items eines Mathematiktests, welche zu zwei Messzeitpunkten bearbeitet wurden. Die Itemschwierigkeiten sind fixiert, so dass angenommen wird, dass alle Verbesserungen bei den Personen stattfinden. Man erkennt in den Balkendiagrammen, dass die Personen sich zwischen dem Messzeitpunkten 1 und 5 verbessert haben. Die einzelnen Itemparameter reichen von -2 zu 2 auf der Logitskala und decken zum ersten Messzeitpunkt die meisten

Personen ab. Zum fünften Messzeitpunkt gibt es aber einige Personen, welche eine höhere Personenfähigkeit haben und somit nicht mehr von den Items gemessen werden können. Der Test benötigt daher mehrere schwere Items und weniger leichte Items, um den Lernverlauf auch zu einem späteren Zeitpunkt noch reliabel zu messen.

Abbildung 10
Person (MZP1 und 5) – Item (MZP1) – Map der Dimension curriculare Aufgaben zum Vergleich der Kompetenzverteilung zu den beiden Messzeitpunkten



Arbeitsauftrag:
Was ist ein Messfehler? Warum ist dieser relevant und inwieweit beeinflusst ein Messfehler die Messung? Was ist der Unterschied zwischen der KTT und der PTT?

Weitere Informationen:
Schurig, M. & Gebhardt, M. (2022). Theoretische Grundlagen von Messungen und Tests. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung.* *Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 233-245). Universität Regensburg.

Arbeitsauftrag zum Text von Schurig und Gebhardt (2022):
Warum ist die KTT besonders empfindlich für systematische Messfehler?
Warum lassen sich die Axiome der KTT in ihrem eigenen Rahmen kaum prüfen?
Wieso soll man Messfehler anhand fachspezifischer Interpretationskulturen prüfen?

Zusätzliche Informationen:
https://www.rasch.org/rasch.htm

4.8 Gütekriterien

Lienert und Raatz definieren einen Test folgendermaßen:

„Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung" (Lienert & Raatz, 1998, S. 1).

Auch wenn es hierzu bereits viel Forschung gibt, warnen bereits Lienert und Ratz, dass in der diagnostischen Praxis Verfahren selbst "zusammengebastelt" (Lienert & Raatz, 1998, S. 5) werden und so nicht sicher ist, ob solche Verfahren den Gütekriterien entsprechen. Aus der Sicht der psychologischen Praxis definieren Lienert und Raatz (1998) als Hauptkriterien die Objektivität, die Reliabilität und die Validität sowie die Nebenkriterien Normierung, Vergleichbarkeit, Ökonomie und Nutzen. Kubinger (2009) betont dagegen Aspekte aus der IRT und sieht neben der Objektivität, der Reliabilität und der Validität, als weitere

Kriterien die Eichung, Skalierung, Ökonomie, Nützlichkeit, Zumutbarkeit, (Un-)Verfälschbarkeit und Fairness an.
Generell geht es bei Betrachtung der Gütekriterien um die Frage, wie möglichst genau und ohne Fehler das gewünschte latente Konstrukt gemessen werden kann. Die Gütekriterien nach Kubinger (2009) werden in Tabelle 6 kurz dargestellt.

Tabelle 6
Gütekriterien nach Kubinger (2009)

Objektivität:	Das Testergebnis ist unabhängig von dem/der Testleiter:in und dem Ort sowie in Durchführung, Interpretation und Auswertung. Es kommt für dieselbe Person theoretisch zum selben Ergebnis und ist daher möglichst standardisiert (Testleiterunabhängigkeit, Verrechnungssicherheit, Interpretationseindeutigkeit).
Reliabilität:	Reliabilität (Zuverlässigkeit) beschreibt den Grad der Genauigkeit der Messung eines latenten Konstrukts von einem Instrument, ohne jedoch über die Gültigkeit eine Aussage zu treffen. So sollte ein Test bei Wiederholung oder in paralleler Form ähnliche Ergebnisse haben (Retest- oder Paralleltestreliabilität). Die Gesamtwerte der Tests sollten hoch korrelieren. Eine andere Form ist die interne Konsistenz, bei welcher die einzelnen Items mit dem Gesamtwert korrelieren sollten (beispielsweise der Cronbachs Alpha).
Validität:	Ein valides Instrument misst das latente Konstrukt, welches es zu messen behauptet. So kann man die inhaltliche Güte von Expert:innen prüfen lassen oder auch der Testperson selbst kann augenscheinlich klar sein, was nun gemessen werden soll. Ebenso lässt sich die Konstruktvalidität prüfen, indem beispielsweise die Ladungsstruktur auf latenter Ebene der theoretischen Struktur des Konstrukts entspricht. Eine Kriteriumsvalidität besteht dann, wenn der Test möglichst hoch mit einem als relevant erachteten Außenkriterium übereinstimmt.
Eichung oder Normierung:	Sobald die ersten drei Gütekriterien erfüllt sind, ist es häufig das Ziel den Testwert einer Person mit einer repräsentativen Vergleichsgruppe (Alter, Schulstufe, Berufsgruppe) zu vergleichen. Hierfür wird eine größere repräsentative Erhebung benötigt. Die Ergebnisse der Erhebung werden transformiert und sind als Normwerte (IQ-Werte) für den Vergleich nach Alter oder Klassenstufe sortiert.
Skalierung:	Die Verrechnungs- und Auswertungsvorschriften des Tests (theoretisches Modell) entsprechen auch dem gemessenen empirischen Modell und die daraus resultierenden einzelnen Testwerte bilden die empirischen Verhaltensrelationen ab. Dies wird verletzt, wenn beispielsweise bei Klausuren (Schulaufgaben) nicht alle Aufgabenschwierigkeiten passend zur empirischen Messung vorliegen und somit entweder zu viele leichte oder schwere Aufgaben verrechnet werden.
Ökonomie:	Sie ist der diagnostische Informationsgewinn in Bezug zu den eingesetzten Ressourcen (Zeit, Kosten, Arbeitsaufwand) für den/die Diagnostiker:in.
Nützlichkeit:	Der Einsatz des Tests erzeugt für die pädagogischen Entscheidungen mehr Nutzen, so dass sich die Anstrengung und Nachteile für die Erhebung bei der Person rechtfertigen.
Zumutbarkeit:	Die Testdurchführung ist für die betroffene Person zumutbar und die Testung selbst wird möglichst schonend durchgeführt.
Unverfälschbarkeit:	Das Testergebnis sollte von der Testperson möglichst nicht bewusst oder unbewusst verfälscht werden können.
Fairness:	Der Test enthält keine systemischen Diskriminierungen oder Nachteile für bestimmte Gruppen und misst alle Personen der Zielgruppe ähnlich fair.

Arbeitsauftrag:
Beschreiben Sie die Gütekriterien in eigenen Worten. Beantworten Sie die Frage:
Wozu werden Gütekriterien in der pädagogischen Praxis benötigt und wann verwenden Anwender:innen die Gütekriterien bewusst oder unbewusst?

Während die Kriterien Objektivität, Reliabilität und Validität bei vielen Tests, wie beispielsweise Schulleistungstests, zutreffen und erfüllt sind, sind die weiteren Kriterien nicht für alle Personen erfüllt. Meist werden Tests für eine bestimmte Zielgruppe entwickelt und messen für einen bestimmten Bereich sehr genau. Für Gruppen, welche definitionsgemäß eher am Rand einer Verteilung sind, treffen insbesondere die Skalierung und auch die Fairness nicht immer zu. Beispielsweise ist ein Definitionskriterium für Kinder mit einer geistigen Behinderung ein niedriger IQ, welcher 2 Standardabweichungen unter dem Mittelwert der Altersnorm liegen soll. Die meisten Intelligenztests sind jedoch für diese Gruppe an Kindern nicht entworfen worden und messen daher nicht unbedingt fair (Ogata, 2019). Eine noch größere Herausforderung ist die Intelligenzmessung bei Kindern mit Sehbeeinträchtigung (Capovilla & Kober, 2019). Aber auch Kinder mit einer anderen Erstsprache (Muttersprache) können insbesondere in Tests mit Bezug auf Sprache, Lesen oder Rechtschreibung benachteiligt sein.

Take Home Message:
Bei jeder Interpretation von Testwerten sollte man sich über die Gütekriterien bewusst sein.
Meist können nicht alle Ansprüche der Gütekriterien im Alltag erfüllt werden. Daher ist es wichtig, solche Limitation in der eigenen Interpretation zu beachten. Insbesondere im sonderpädagogischen Bereich, in welchem Anpassungen und Adaptionen notwendig sein können, ist eine Begründung für das gewählte Vorgehen zu dokumentieren.

Die Gütekriterien sind daher eine entscheidende Richtschnur für die Auswahl an Tests. Dabei ist jedoch zu bedenken, dass nicht jeder Test für alle Personen passt. Jeder Test hat eine bestimmte Zielgruppe, für die er konstruiert wurde und deren Merkmale empirisch geprüft wurden. Viele Tests stammen aus der medizinischen und psychologischen Forschung und sind daher eher für klinische Fragestellungen oder eine umfängliche Statusdiagnostik relevant. Insbesondere im Bereich der pädagogischen Entscheidungen in Förderung und Unterricht gibt es dagegen eher wenig Forschung im deutschsprachigen Bereich im Vergleich zu den USA. Damit verbunden ist auch, dass es weniger Tests für unterschiedliche Einsatzgebiete oder spezifische Gruppen mit besonderen Anforderungen gibt.

Die Gütekriterien werden meist in Bezug auf eine möglichst genaue Messung für eine umfassende Einzeluntersuchung betrachtet. Für die Anwendungsform pädagogischer Handlungen, beispielsweise bei den Einschätzungen zum Lernstand und der Lernentwicklung im schulischen Kontext, spielen insbesondere die Ökonomie und Nützlichkeit eine hervorgehobene Rolle. Ansonsten besteht die Gefahr, dass Tests und standardisierte Verfahren kaum oder gar nicht eingesetzt werden und stattdessen zur Ermittlung des Leistungsstandes ein selbstentworfenes Arbeitsblatt genügt. Aus diesem Grund sind die Gütekriterien je nach Instrument verschieden zu gewichten. Während gravierende pädagogische Entscheidungen auf Basis von mehreren möglichst erprobten und genauen Instrumenten getroffen werden sollten, genügt es für die didaktische Entscheidung Screenings einzusetzen.

Weitere Informationen:
Grabowski, F., Castello, A. & Brodersen, G. (2022) Diagnostische Gütekriterien bei Statustests. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 247-258). Universität Regensburg.

Arbeitsauftrag:
Vergleichen Sie die von Grabowski et al. (2022) beschriebenen Gütekriterien mit den in diesem Text verwendeten Gütekriterien. Wo sind die Gemeinsamkeiten und wo die Unterschiede?

4.9 Messinvarianz und Adaptionen

Damit Tests für alle Kinder interpretiert werden können, müssen die Gütekriterien und insbesondere das Gütekriterium der Skalierung auch bei allen Schüler:innengruppen getestet werden. Dies geschieht, indem bei der Normierung verschiedene Gruppen beachtet und prüft werden, ob die einzelnen Items und ihre Zusammenhänge für alle Schüler:innengruppen ähnlich leicht bzw. schwer sind. So sollten verschiedene Items bei Textaufgaben ähnlich schwer für Kinder mit und ohne Migrationshintergrund oder sonderpädagogischen Förderbedarf sein. Ist dies der Fall sind die Instrumente fair und für die verglichenen Schülergruppen messinvariant. Das bedeutet, dass man dann auch die Werte für die verschiedenen Gruppen gleich interpretieren darf. Umso komplexer Tests sind, umso eher besteht jedoch die Gefahr, dass einzelne Aspekte unfair für bestimmte Gruppen sind. Es kann sein, dass bestimmten Gruppen Vorwissen oder Strategien fehlen, welche zur Bearbeitung der Aufgaben notwendig sind. Dies ist daher im Bereich der Bildungsvergleichenden Studien wie PISA, PRILS und auch dem NEPS ein wichtiger Schwerpunkt.

Take Home Message:
Jeder Test wurde für eine bestimmte Zielgruppe entworfen und für bestimmte Gruppen überprüft. Beachten Sie daher in der Testbeschreibung, ob die Fairness auch für die für Sie relevante Schüler:innengruppe beschrieben ist.

Für die sonderpädagogische Praxis müssen bestimmte Tests adaptiert werden, damit diese überhaupt durchführbar sind. Dies betrifft einerseits die Instruktion vor der Aufgabe und die Beispielaufgaben. Dies ist eher unproblematisch zu sehen, da dadurch sichergestellt wird, dass die einzelne Person das Aufgabenformat verstanden hat. Eine Veränderung der Durchführung ist dagegen kritischer zu sehen. So kann mehr Zeit bei der Durchführung und (standardisierte) Hilfestellungen auch die Testergebnisse beeinflussen. Dies kann notwendig sein, jedoch im Zweifelsfall dazu führen, dass man die Ergebnisse eher als informelles Verfahren interpretieren sollte.

Weitere Informationen:
Renner, G. & Scholz, M. (2022) Fair oder nicht fair, das ist hier die Frage! Die Sicherung der Testfairness als Aufgabe der sonderpädagogischen Diagnostik. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 259-274). Universität Regensburg.

Arbeitsauftrag zum Beitrag von Renner und Scholz (2022):
Was verstehen Renner & Scholz unter Zugangsfertigkeiten?
Was verstehen Renner & Scholz unter Adaptationen?

Exkurs: Zusammenstellungen aus Subtests von verschiedenen Tests
Ebenso gibt es das Vorgehen in der Praxis, dass aus mehreren Tests einzelne Subtests entnommen werden, um diese zu einem neue Gesamttest zu verbinden. Sind die einzelnen Subtests auch als eindimensionaler Test konstruiert und eine eigene Normierung angegeben, dann kann dieses Vorgehen als eine Zusammenstellung von verschiedenen Tests betrachtet werden. Sind die einzelnen Subtests aber nur Teil eines anderen multidimensionalen Tests und liegt dafür keine eigene Norminterpretation vor, sollte man solche Zusammenstellungen aus verschiedenen Tests nur als informelle Tests betrachten.

4.10 Testeffekte

Lernen geschieht in allen Situationen. Auch in Testsituationen kann gelernt werden, selbst wenn nicht die richtige Lösung gezeigt wird. Die Anwendung eines Tests mit Verbindung zum Lernen kann somit Testeffekte aufweisen oder zu abrufbasiertem Lernen (Tempel & Pastötter, 2021) führen. Ebenso kann ein Test vor dem Lernen positive Effekte haben. Für die Schule bedeutet dies, dass die Erfassung eines Lernverlaufs mit mehreren Tests bereits zu einer Verbesserung führen kann. Dies kann sowohl an motivationalen Faktoren als auch an einem verbesserten Informationsabruf auf Seiten der Lernenden liegen.

Die Verbindung von Lernen und Abfragen kennt man aus dem Bereich des Vokabellernens bei einer Fremdsprache. Karpicke und Roediger (2008) fanden für die Merkleistung von Suaheli-Englisch Vokabeln deutlich mehr Lernerfolg, wenn diese wiederholt getestet und bei gescheitertem Abrufversuch wiederholt gelernt wurden, im Vergleich dazu, wenn die Vokabeln nur wiederholt wurden. Generell gilt, dass eine aktive Abfrage bzw. Testung nach einem erfolgten Lernen aus dem Langzeitgedächtnis zu einem besseren Behalten führt als reines Wiederholen. Tempel und Sollich (2023) zeigten in einer Studie mit 36 Kindern einer Förderschule den positiven Effekt eines solchen abbrufbasierten Lernens (Retrieval-based learning).

Mehr Informationen unter:

Tempel, T. & Sollich, S. (2023) Retrieval-based learning in special education. *Journal of Research in Special Educational Needs*, 00, 1– 7. https://doi.org/10.1111/1471-3802.12594

5 Überblick über verschiedene Formen von diagnostischen Instrumenten

Testverfahren sind je nach der pädagogischen Fragestellung unterschiedlich entwickelt. So gibt es Verfahren, welche man als Screenings möglichst schnell und einfach umsetzen kann und andere Verfahren, welche möglichst genau einen Einzelfall messen sollen. Die Frage, welchen Umfang die Instrumente haben, hat dabei auch mit der Frage von Sensitivität und Spezifität zu tun.

- Sensitivität misst den Anteil der in der Wirklichkeit mit diesem Merkmal positiven Personen, welche durch einen Test als positiv identifiziert wurden (Positiv/Positiv).
- Spezifität misst den Anteil der in der Wirklichkeit mit diesem Merkmal negativen Personen, welche durch einen Test als negativ identifiziert wurden (Negativ/Negativ).

Für den/die Pädagog:in stellt sich die Frage, ob ein kurzes Instrument genügt, um einen ersten Einblick zu gewinnen, einen Anfangsverdacht zu bestätigen oder eine präventive pädagogische Entscheidung zu treffen. Wenn dem so ist, kann ein ökonomisches, aber möglichst sensitives Instrument verwendet werden. Mit solchen Verfahren werden Personen identifiziert, welche ein besonderes Risiko haben. Hierbei wird in Kauf genommen, dass auch Personen identifiziert werden, welche bei genauerer Betrachtung oder einer umfangreicheren Testung ein solches Risiko nicht hätten (Falsch-Positiv). Für das präventive Arbeiten sind solch sensible Instrumente als Screenings zu bevorzugen.
Besteht aber die Gefahr einer Kategorisierung oder auch negativer Folgen durch eine Klassifizierung, muss auch die Spezifität beachtet werden. In der pädagogischen Praxis bedeutet dies, dass man dann umfangreichere Verfahren zur Statusdiagnostik verwenden sollte und weitere Beobachtungen als Information für die Argumentation und Interpretation beachtet. Umso mehr reliable Informationen verwendet werden, umso sicherer ist die daraus abgeleitete Entscheidung. Auf der anderen Seite muss sich eine umfangreiche Diagnostik auch für den weiteren Verlauf besonders für die betroffene Person als hilfreich erweisen und daher auch mit pädagogischen Entscheidungen sowie möglichen Maßnahmen einhergehen. Dies ist beispielsweise im Schulsystem jedoch nicht immer der Fall. „Eine umso mehr, desto besser"-Vorgehensweise ist daher nicht zu empfehlen.

Nach Hasbrouck und Tindal (2006) können vier Kategorien von diagnostischen Instrumenten je nach Einsatzgebiet unterschieden werden:

- **Screenings** messen zu einem Messzeitpunkt einen Teil eines Lernbereichs, um möglichst gut Kinder mit Schwierigkeiten zu entdecken.
- **Schulleistungstests für Gutachten (Statustests)** sind normierte Tests, welche man zu jederzeit im Schuljahr durchführen kann, um *ein Kind* zu einem Messzeitpunkt in einem Kompetenzbereich möglichst umfänglich zu testen.
- **Lernverlaufsdiagnostik** besteht aus mindestens drei meist kürzeren Tests, welche in einem kontinuierlichen (wöchentlichen oder monatlichen) Abstand erhoben werden, um den Lernverlauf eines Kindes über die Zeit darzustellen.
- **Kriteriumsbezogene Tests** sind umfassende Schulleistungstests, um zu bestimmen, ob die Schüler:innen das Klassenziel oder das anzuwendende Kriterium erreicht haben.

Tests werden für unterschiedliche Einsatzgebiete entwickelt. Zwar gelten kommerzielle Tests als Goldstandard, dies ist aber insbesondere für Einsatzgebiete außerhalb der Individualdiagnostik für Gutachten mehr als fraglich. Insbesondere für pädagogische Entscheidungen im Unterricht im Bereich des classroombased Assessment gibt es gute Forschungsinstrumente, welche die Wirkung von Interventionen besser erfassen. Daher bemüht sich dieses Buch solche freien Instrumente bevorzugt zu verlinken.

Arbeitsauftrag:
Die Informationsdatenbank diagnostischer Verfahren der PH Ludwigsburg - Dia-Inform stellt auf wenigen Seiten umfangreiche Überblicke über die Anwendung und Durchführung von verschiedenen Testverfahren zusammen und bewertet diese Verfahren. Hierbei wird auch der WISC-V von Renner und Schroeder (2018) analysiert. Lesen Sie insbesondere die Bewertung des Tests durch und beschreiben Sie in eigenen Worten, für wen sich dieser Test besonders eignet und für welche Gruppen eine faire Testung unter Umständen schwierig wäre.

Weitere Informationen:
Buchwald, K., Ebenbeck, N. & Gebhardt, M. (2022). Screenings, Status- und adaptive Tests in der schulischen Diagnostik. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 305-311). Universität Regensburg.

5.1 Power- oder Speedtest

Der **Powertest** fängt mit leichten Aufgaben an und wird immer schwieriger. Während die Anzahl der Aufgaben begrenzt ist, ist die dafür verfügbare Testzeit theoretisch unbegrenzt. Besteht ein Test beispielsweise aus 25 Aufgaben, dürfen alle Personen so lange brauchen, wie sie es selbst für nötig erachten. Sobald die Person drei nacheinander folgende Aufgaben nicht lösen kann, darf man abbrechen. Im optimalen Fall wird mindestens eine Aufgabe und maximal 24 Aufgaben richtig beantwortet. Denn wenn eine Person keine Items richtig beantwortet hat, kann die Fähigkeit der Person unter dem Testbereich liegen und bei der Beantwortung aller Items über dem Niveau des Tests. In beiden Fällen liegt die Person außerhalb des Testbereichs. Für den Alltag in der Schule ist ein Test ohne Zeitbegrenzung schwierig durchzuführen. Insbesondere als Gruppentest hat man das Problem, dass ein Teil der Schüler:innen nach wenigen Minuten und andere Schüler:innen erst nach einem längeren Zeitraum die Aufgaben beenden.

Take Home Message:
Powertests haben kein Zeitlimit und werden mit jeder Aufgabe schwieriger.

Aus diesem Grund haben Tests insbesondere als Computertests eine Zeitbegrenzung. Insbesondere **kurze Tests von ein- bis fünfminütigen** Screenings oder Lernverlaufsdiagnostiktests wurden so entwickelt, dass man diese einfacher in den Unterrichtsalltag, während einer Freiarbeitsphase integrieren kann. Diese Tests haben ein Zeitlimit und dies bedeutet, dass die Bearbeitungsgeschwindigkeit neben der Lösungswahrscheinlichkeit eine Rolle spielt, um einen hohen Wert zu erreichen. Tests, welche als **Speedtests** konstruiert sind, enthalten viele ähnlich schwere Items, die insgesamt sehr leicht gelöst werden können. So misst man insbesondere die Bearbeitungsgeschwindigkeit. Tests zur Messung der Leseflüssigkeit sind meist als einminütige Speedtests konstruiert. Schüler:innen in den ersten Jahrgangstufen sollen möglichst wenig Lesefehler machen und zudem möglichst schnell lesen. Daher ist es wichtig, einen Test auszuwählen, welcher von dem/r individuellen Schüler:in möglichst fehlerfrei bearbeitet werden kann. So kann in einem ersten Schritt als Leseflüssigkeit ein Lesetest basierend auf Silben gewählt werden. Bei ausreichend hoher Lösungswahrscheinlichkeit von über 90% kann auf ein höheres Niveau (Schwierigkeit) oder einen Lesetest mit Wörtern oder Pseudowörtern gewechselt werden.

Take Home Message:
Speedtests enthalten viele leichte und/oder gleichschwere Aufgaben mit Zeitlimit, so dass nicht alle Aufgaben von den Personen bearbeitet werden können. Die Messung der Geschwindigkeit steht im Fokus.

Tests, welche beide Formen mischen, werden speeded Tests genannt. Häufig enthalten diese Tests Items mit verschiedenen Schwierigkeiten und eine Zeitbegrenzung. In solchen Fällen muss geprüft werden, ob die Anzahl der gelösten Aufgaben eher an der Bearbeitungsgeschwindigkeit oder der Fähigkeit der Person lag. Daher sollte bei der Auswertung die Anzahl der gelösten Aufgaben als auch die Lösungswahrscheinlichkeit (Anteil der gelösten Aufgaben an allen bearbeiteten Aufgaben) beachtet werden.

Ebenso sollte man auch nicht mehrere Tests an einem Termin hintereinander durchführen. Insbesondere bei dem zuletzt durchgeführten Test ist vermutlich weder die Konzentration noch die Motivation ausreichend, um die normale oder reguläre Leistung der Person insbesondere bei Speedtests zu zeigen. Ausreichende Pausen oder verschiedene Termine sind daher notwendig. Insbesondere bei Kindern mit sonderpädagogischem Förderbedarf ist dies zu berücksichtigen. Auch wenn dieser Hinweis dem/der Leser:in redundant erscheint, ist dies ein immer wieder anzutreffender Fehler in der Praxis.

Arbeitsauftrag:
Führen sie einen Leseflüssigkeitstest mit Pseudowörtern mit ihrem/r Banknachbar:in entweder mit der Paper Pencil Variante oder in Levumi.de durch.

SiL-Levumi - Tests der Leseflüssigkeit zur Lernverlaufsdiagnostik - "Silben lesen" der Onlineplattform www.levumi.de
WoL- Levumi-Tests der Leseflüssigkeit zur Lernverlaufsdiagnostik - „Wörter lesen" der Onlineplattform www.levumi.de.
PseuWoL-Levumi - Tests der Leseflüssigkeit zur Lernverlaufsdiagnostik „Pseudowörter lesen" der Onlineplattform www.levumi.de.

Beschreiben Sie, ob die Tests von Levumi Powertests und/oder Speedtests sind. Begründen Sie Ihre Entscheidung.

5.2 Klassenstufen- oder konstruktbasierte Schulleistungstests

Tests konstruiert man je nach Anforderung und Anwendung unterschiedlich mit dem Ziel, für die pädagogische Fragestellung eine möglichst reliable Aussage zu erhalten. Nachdem insbesondere die Frage der Selektion von Bedeutung ist, werden die meisten Schulleistungstests nach Anforderung einer Klassenstufe konstruiert.

Die pädagogische Fragestellung ist somit, ob das Kind alle Anforderungen der Klassenstufe erfüllt oder nicht. Dies ist eine systemische Blickrichtung, welche sich nach den Bedürfnissen der Lehrkräfte richtet. Dass der Lernbereich Deutsch oder Mathematik aus mehreren Kompetenzbereichen und Fertigkeiten besteht, wird insofern berücksichtigt, als für die Bildung eines intervallskalierten Wertes mehrere Subtests verwendet werden. Demat 2+ besteht beispielsweise aus den zehn Subtests Zahleneigenschaften, Längenvergleich, Addition und Subtraktion, Verdoppeln und Halbieren, Division, Rechnen mit Geld, Sachaufgaben und Geometrie. Diese Gesamtmischung eines **Schulleistungstests** ergibt das mathematische Profil in der zweiten Klasse nach dem Lehrplan. Für eine grobe Übersicht kann diese unterschiedlichen Bereiche zusammensetzt werden. Man muss sich jedoch bewusst sein, dass die Kompetenzen in Geometrie im Gesamtergebnis nur einen kleinen Teil ausmachen und an sich völlig andere Konstrukte regeln und auch andere Lernbereiche abdecken als beispielsweise die Subtests der Grundrechenarten.

Anstatt einen Test nach den Inhalten des Lehrplans zu entwerfen, könnte man das mathematische Kompetenzprofil auch durch einzelne Tests, beispielsweise zum Zahlenraum, zu den Grundrechenarten, zu Sachaufgaben und zu Geometrie bilden. Solche einzelnen Tests werden nach den latenten Konstrukten hinter den Tests gebildet und können auch für sich aus einer oder mehreren Dimensionen bestehen. So kann ein Grundrechentest beispielsweise aus Subtests zum Addieren, Subtrahieren, Multiplizieren oder Dividieren bestehen. Umso spezifischer das latente Konstrukt ist, desto näher ist der Test an einem sogenannten **Skilltest,** bei welchem die Person nur bestimmte Fähigkeiten/Fertigkeiten oder eine bestimmte Strategie zur Lösung anwenden braucht (Addieren im 10er Bereich). Skilltests bzw. **konstruktbasierte Tests** messen eine Kompetenz (z. B. die Lesekompetenz) bzw. ein Konstrukt nach theoretischen Erkenntnissen. Auch hier gibt es sowohl kleinere Tests, die enge Kompetenzbereiche testen (z. B. das Lesen von bestimmten Silben mit wenigen Buchstaben), als auch Tests, die breite Kompetenzbereiche testen (z. B. Addieren und Subtrahieren zwischen 1 und 100). Die Auswahl solcher konstruktbasierter Tests ist individueller, da nicht der generelle Bezugsrahmen der Jahrgangsstufe, sondern die Kompetenz des Kindes eingeschätzt werden muss. Konstruktbasierte Tests beziehen sich nicht auf eine genaue Personengruppe, Altersgruppe oder Jahrgangsstufe, sondern auf Kriterien, welche bei der Konstruktion festgelegt worden sind.

Take Home Message:
Skilltests messen einen eng definierten Skill bzw. eine Fertigkeit, welche meist gut trainierbar ist.

Anstatt das Lesen nach den Inhalten der ersten oder zweiten Klasse zu trennen, misst man mit **konstruktbasierten Lesetests** entweder die Leseflüssigkeit oder das sinnkonstruierende Lesen. Die Leseflüssigkeit wird beim SiL-Levumi (Jungjohann et al., 2019) mit einem einminütigen Test gemessen. Es stehen mehrere Tests in unterschiedlichen Niveaustufen mit verschiedenen Buchstaben nach dem Aufbau des Kieler Leseaufbaus zur Verfügung. Alle Silben sind aus dem Grundwortschatz und pro Test ähnlich schwierig.

Da nicht nur in inklusiven, sondern auch in Klassen der Grund- und Sekundarstufe die Leistungen sehr heterogen sind, kann es sinnvoll sein, einen konstruktbasierten Test für eine Teilkompetenz zu wählen. Dieser misst in diesem Bereich meist genauer, aber unter Umständen auch breiter, da er mehr Aufgaben enthalten kann als curriculumsbasierte Schulleistungstests mit einer Klassenstufe. Mittlerweile gibt es daher auch Tests, welche die Inhalte mehrere Klassenstufen enthalten.

5.3 Regelbasierte Itemkonstruktion

Damit die Ergebnisse von Tests interpretierbar sind, werden Tests nach **regelbasierter Itemkonstruktion** entwickelt**.** Meist werden diese Regeln nur implizit geprüft und den Testentwickelnden aus, dass der Test an sich empirisch geprüft wurde. Man kann jedoch auch die eigenen Konstruktionsregeln prüfen. Dieses Vorgehen ist notwendig, um beispielsweise den Computer selbst nach einem regelgeleiteten Ansatz weitere Items entwerfen zu lassen. Dies wird Automatic Item Generation (AIG, Automatische Itemgenerierung) genannt.

Für jeden Test ist ein zentrales Element, dass Regeln aufgestellt werden, warum ein Item für die angenommene Zielgruppe schwierig oder leicht sein könnte. Solche Regeln werden Difficulty generating item characteristics (DGICs) genannt. Mittels des linear logistischen Test-Modells (LLTM; Fischer, 1973), einer linearen Erweiterung des Rasch-Modells (RM; Rasch, 1980), lässt sich bei gegebener Gültigkeit des RM die Zusammensetzung der Itemschwierigkeiten als Funktion der DGICs abbilden. Jedes einbezogene Merkmal erhöht die Schwierigkeit der Aufgabe um einen spezifischen Wert. Die Gesamtschwierigkeit einer Aufgabe ergibt sich dann aus der gewichteten Summe der einzelnen Werte. Die einzelnen

Gewichte werden vorab auf Basis theoretischer Annahmen in einer Designmatrix festgelegt (Wilbert, 2014).

Beispiel:
ZsM100-Levumi N3 Tests zu Rechenoperationen als Lernverlaufsdiagnostik. Zahlen subtrahieren (Zahlenraum 100) der Onlineplattform www.levumi.de

Für den Test **Zahlen subtrahieren (Zahlenraum 1 – 100)** wurden mehrere Aufgabenmerkmale bei der Konstruktion berücksichtigt. Eine Aufgabe ist beispielsweise schwieriger, wenn sie einen Zehnerübergang beinhaltet. Ebenso wurde berücksichtigt, ob eine Ziffernanzahl im zweiten Term einstellig oder zweistellig und der Minuend oder der Subtrahend als Zehner vorkommt. Je nach Gewichtung dieser Merkmale sind die Aufgaben unterschiedlich schwer und werden in unterschiedliche Pools von Aufgabenkategorien hinterlegt, was in der Tabelle 7 zu erkennen ist.

Take Home Message:
Regelbasierte Tests werden nach Regeln konstruiert, welche die Schwierigkeiten definieren. Die Ergebnisse dieser Tests sind dann leichter zu interpretieren, wenn die Regeln nachvollziehbar beschrieben werden.

Tabelle 7
Kategorien für regelbasierte Aufgaben des ZsM100-Levumi N3 Tests

Aufgaben-kategorie	Aufgaben-beispiel	Notwendigkeit eines Zehner-übergangs	Ziffernanzahl im zweiten Term	Minuend als Zehner	Subtrahend als Zehner
1	67 – 6	nein	einstellig	nein	nein
2	60 – 40	nein	zweistellig	ja	ja
3	47 – 20	nein	zweistellig	nein	ja
4	95 – 53	nein	zweistellig	nein	nein
5	30 – 6	ja	einstellig	ja	nein
6	23 – 7	ja	einstellig	nein	nein
7	80 – 34	ja	zweistellig	ja	nein
8	57 – 29	ja	zweistellig	nein	nein

Bei der Zusammenstellung der einzelnen Tests werden aus den acht Aufgabenkategorien gleich viele Items gezogen, um möglichst ähnlich schwere parallele Tests zu konstruieren.

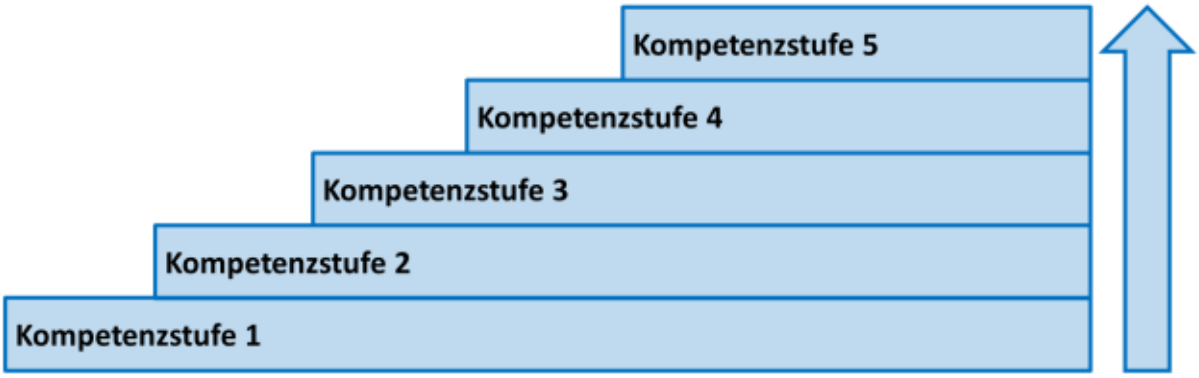
Arbeitsauftrag:
Lesen Sie die Testbeschreibung zu den Platzhalteraufgaben und beschreiben Sie die schwierigkeitsgenerierenden Merkmale des Tests.

5.4 Kriterienorientierte Tests

Kriterienorientierte Tests basieren meist auf einem theoretischen Stufenmodell, häufig mit Lehrplan- oder Kompetenzbezug, welches als Außenkriterium dient. Nach diesem sind geregelt, welche konkreten Abstufungen in den Schwierigkeiten der Items begründet sind. Daher bestehen solche Tests aus unterschiedlich schweren Aufgaben und basieren auf der Idee der Powertests. Wie in Abbildung 11 dargestellt sind die Aufgaben meist als Teil einer Kompetenzstufe konstruiert und können daher direkte Empfehlungen für den Unterricht geben. Aus diesem Grund sind kriterienorientierte Tests meist mit Förderempfehlungen und konkreten Fördermaterial verbunden. Ein Beispiel für einen kriterialen freiverfügbaren Test ist der Test Lost in Math – LoMa von Ehlert et al. (2022) oder Ilea – Individuelle Lernstandsanalysen von Prengel und Liebers (2005).

Take Home Message:
Kriterienorientierte Tests sind eng mit dem Lehrplan oder einem Kompetenzstufenmodell verbunden. Es werden Werte definiert, welche einen erreichten Kompetenzbereich definieren.

Abbildung 11
Vereinfachte Darstellung eines Kompetenzstufenmodells



Nicht alle Tests haben ein explizit ausformuliertes Stufenmodell. Viele Tests verwenden jedoch Kompetenzmodelle und können dadurch auch in einer oder mehreren Stufen interpretiert werden. So verwenden beispielsweise mehrere Tests von Levumi verschiedene Schwierigkeitsstufen, welche auch als Kompetenzstufen betrachtet werden können. Ebenso gibt es auf Levumi meist mehrere Niveaustufen einer Testform. Zusammengenommen können diese auch als ein Test mit verschiedenen Kompetenzstufen betrachtet werden.

Arbeitsauftrag:
Gehen Sie auf die Internetseiten von LoMa oder Ilea und lesen Sie sich die Lehrkrafthefte bzw. das Manual durch. Fassen Sie in eigenen Worten zusammen, wie der Test angewendet wird und ob und wie dieser Test mit Förderschritten verbunden ist.

5.5 Screenings

Screenings sind kurze ökonomische Tests, welche reliabel Kinder mit Schwierigkeiten identifizieren möchten und einen schnellen Überblick über einen spezifischen Leistungsstand einer Person oder einer Klasse geben. Screenings messen möglichst ökonomisch zu einem Messzeitpunkt einen Teil eines Lernbereichs, welcher für das Erlernen einer Kompetenz ausschlaggebend ist. Aus den Ergebnissen kann der Ist-Stand eines Kindes im jeweiligen Lernbereich abgeleitet werden, weshalb der Einsatz zu Beginn des Schuljahres empfohlen wird (Hasbrouck & Tindal, 2006). Lehrkräfte verwenden Screenings, um einen schnellen und einfachen Überblick über den Leistungsstand einer Gruppe oder einer Person zu erhalten. Die ökonomische Handhabung steht im Vordergrund, damit die Tests z. B. einfach und schnell in inklusiven Klassen durchgeführt werden können. So können Kinder mit Schwierigkeiten frühzeitig erkannt und zusätzliche Förderung zum Klassenunterricht eingeleitet werden (Gebhardt, 2021; Hartung et al., 2021). Screenings dienen somit in erster Linie der didaktischen Entscheidungsfindung, um insbesondere Risikokindern präventiv passendere und bessere Förderung zukommen zu lassen.

Take Home Message:
Screenings sind kurze Tests, welche einen Überblick oder eine erste Einschätzung über einen Lernbereich geben.

Ein kostenfreies Screening für den Bereich Mathematik ist beispielsweise die Reihe Mathes (Sikora & Voß, 2016), welches von der Einschulung bis zur 4. Klasse entworfen und geprüft ist. Das Verfahren dauert je nach Klassenstufe zwischen 30 und 45 Minuten. Da das Screening mehrere Aufgabengruppen und somit Kompetenzbereiche umfasst, dauert es sehr lang. Ebenso kann die Screeningdurchführung auch nur wenige Minuten benötigen. In diesem Fall messen die Screenings jedoch ein weniger breites Testprofil. Screenings selbst wollen nur einen Einblick geben und sind daher kürzer sowie schneller durchführbar als umfangreiche Schulleistungstests. Für eine umfassende Einzelfalldiagnostik und die Feststellung eines sonderpädagogischen Förderbedarfs im Rahmen eines sonderpädagogischen Gutachtens sind Screenings daher nicht geeignet.

Arbeitsauftrag:
Wählen Sie ein Screening von Levumi oder Lernlinien aus und erörtern Sie anhand dessen, für welches Einsatzgebiet dieses Instrument in der inklusiven Schule passend würde.

Weitere Informationen:
Ebenbeck, N., Jungjohann, J. & Gebhardt, M. (2022). *Testbeschreibung des Lesescreenings LES-IN für dritte inklusive Klassen.* Universität Regensburg. https://doi.org/10.5283/epub.53204

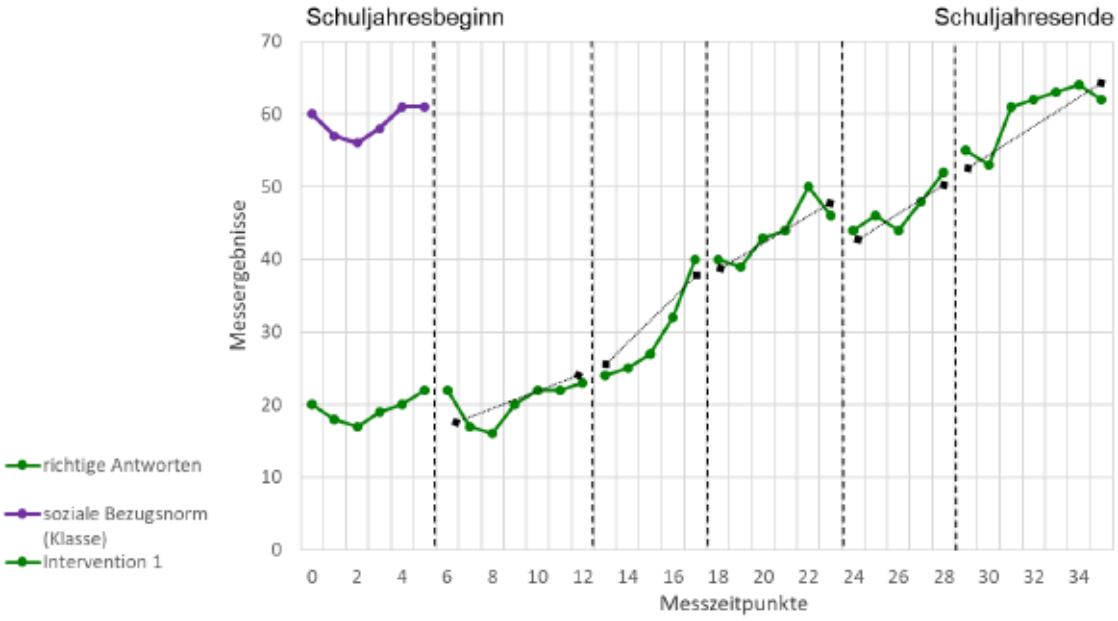
Sikora, S. & Voß, S. (2016). *Mathes 0. Schuleingangstest zur Erfassung arithmetischer Vorläuferfähigkeiten*. http://www.lernverlaufsdiagnostik.de

5.6 Verlaufsdiagnostiktests

Verlaufsdiagnostik ist eine quantitative Form der formativen Diagnostik und misst meist begleitend zu Unterricht und Förderung den Fortschritt im Lernen oder Verhalten (Voß & Gebhardt, 2017). Im **Bereich Verhalten** können einzelne Items oder mehrere Items das Verhaltensziel abbilden und dieselbe Person bewertet anhand der Items den gelegten Bereich des Verhaltens in einem festgelegten Zeitraum. Die Schülerin Susi arbeitet mit der Lehrkraft beispielsweise am Verhaltensziel, in der ersten Stunde ruhig am Platz zu sitzen. Daher wählt die Lehrkraft aus dem Direct Behavior Rating PUTSIE (Schurig et al., 2020) das Item „Arbeitet ruhig am Platz" aus und bewertet jeden Tag in der ersten Stunde, ob Susi ihr Ziel nie (= 1) oder immer (= 7) erreicht. Ihre Bewertung meldet sie Susi individuell zurück und beide überlegen, wie man das Ziel noch besser erreichen kann. Ist das Ziel erreicht, kann man auch jedes andere gut beobachtbare Verhalten in Form eines Items operationalisieren oder aus bestehenden Instrumenten auswählen.
Die Einstufung des Verhaltens ist bei derselben Person in einem kurzen Zeitverlauf relativ stabil (Gebhardt et al. 2020). Ob das eingeschätzte Verhalten, auch das objektive Verhalten ist, ist fraglich. Aus diesem Grund erscheint es sinnvoll, mehrere Personen zu mehreren Zeitpunkten in die Bewertung des Verhaltens einzubeziehen. Dann kann man die Übereinstimmungen der Bewerter:innen über die Zeit vergleichen (Casale et al., 2017).

Die Messung von **Lernverläufen** ist komplexer, da nicht dieselben Items pro Testung verwendet werden können. Trotzdem muss der Test aber zu vielen Terminen fair und sensibel messen. Für die Praxis praktikable und einfach anwendbare Tests zu konstruieren, welche psychometrisch jedoch reliabel messen, ist sehr aufwendig. Am besten gelingt dies anhand von Computertests, welche die Durchführung und Auswertung übernehmen. Hier ist es möglich, dass jede Person anhand von Ziehungen aus einem Pool an Items ähnlich schwere Messungen für jeden Messzeitpunkt bekommt und somit theoretisch unendlich viele Messzeitpunkte möglich sind (Gebhardt et al., 2021).

Abbildung 12
Vereinfachte Darstellung des Lernverlaufsgraphen

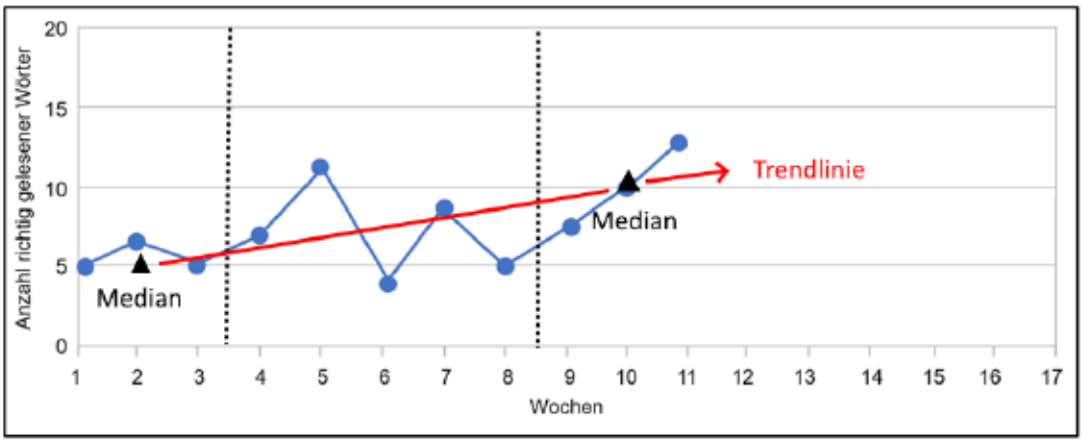


In Abbildung 12 ist ein exemplarischer Lernverlaufsgraph anhand eines einminütigen Lesetests dargestellt. Grüne Punkte stehen für die Anzahl der richtig gelösten Wörter zum jeweiligen Testzeitpunkt, während die lila Punkte die kumulierten Ergebnisse der Klassenkamerad:innen darstellen. Ebenso sind in der Abbildung 12 mit den gestrichelten Linien von der Lehrkraft verschiedene Interventionsphasen eingezeichnet. Zudem sind mittels eines Lineals händische Trendlinien geschätzt worden. Dies ermöglicht eine grobe Einschätzung des Verlaufs über die Zeit, wenn die Reliabilität des Instruments und der Messung sowie die Anzahl der Messungen dies zulassen.

Lernverlaufsdiagnostik sollte regelmäßig durchgeführt werden. Sie dauert, um eine praktische Umsetzung im Unterricht zu ermöglichen, meist nur wenige Minuten. Aus diesem Grund sind viele Tests, insbesondere die, die am Computer durchgeführt werden, als Speedtests mit fester Zeitvorgabe konstruiert.

Für die Interpretation von Lernverlaufsgraphen sind insbesondere Trendlinien wichtig. Diese können bei Computerauswertungen anhand von Regressionsverfahren gebildet werden. Für den einfachen Gebrauch in der analogen Praxis hat sich die Methode „Tukey Tri Split" durchgesetzt (Jungjohann et al., 2022). Der bestehende Lernverlaufsgraph wird in drei gleich große Abschnitte unterteilt (siehe Abbildung 13). Kann der Graph nicht in drei gleich große Abschnitte unterteilt werden, wird die Unterteilung so vorgenommen, dass das erste und dritte Segment gleich lang und der Mittelteil entsprechend länger ist. Für den ersten und den letzten Abschnitt wird anschließend der Median bestimmt, da dieser weniger anfällig für Ausreißer ist als der Mittelwert. Der Median wird jeweils im Lernverlaufsgraphen markiert. Anschließend werden die beiden Medianwerte verbunden und daraus eine Trendlinie erstellt. Die aus dem Tukey Tri Split resultierende Trendlinie kann zur Abschätzung der zukünftigen Lernentwicklung verwendet werden, wenn von einem weiterhin gleichbleibenden Unterricht bzw. Fördermaßnahmen ausgegangen wird. Ansonsten sind die weiteren Auswertungsmethoden im Sinne der kontrollierten Einzelfallforschung empfohlen (siehe Kapitel 8).

Abbildung 13
Vereinfachter Lernverlaufsgraph mit Trendlinie



Video:
Lernverlaufsdiagnostik (15:48)

Weitere Informationen:
Aktuelle Übersicht zur LVD auf der Homepage des Lehrstuhls

Gebhardt, M., Jungjohann, J. & Schurig, M. (2021). *Lernverlaufsdiagnostik im förderorientierten Unterricht. Eine Einführung in Anwendung, Testkonstruktion und Instrumente.* Reinhardt Verlag.

Blumenthal, S. (2022). Lernverlaufsdiagnostik. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 633-648). Universität Regensburg.

Breucker, T. & Kuhl, J. (2022). Formative Diagnostik in der deutschen Sonderpädagogik. *In* M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 619-632). Universität Regensburg.

Arbeitsauftrag:
Berechnen Sie nach der Methode des „Tukey Tri Split" die Medianwerte der LVD und zeichnen Sie diese in einem Graph ein. Die Ergebnisse der LVD waren in chronologischer Reihenfolge:

Kind A: 3, 6, 5, 8, 12, 15, 15, 18, 15
Kind B: 1, 3, 5, 9, 6, 8, 9, 10, 10
Kind C: 5, 2, 1, 4, 8, 9, 8, 10, 9, 9

5.7 Adaptive Tests

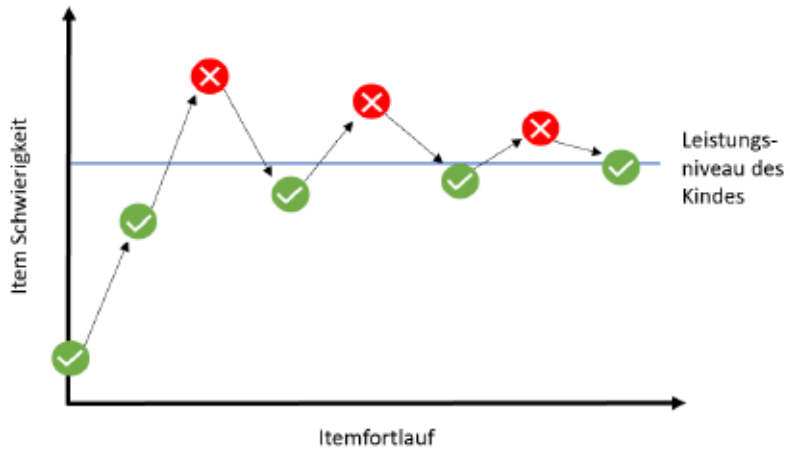
Reguläre Powertests fangen mit den einfachsten Items (Aufgaben) an und werden im Laufe der Testdurchführung immer schwieriger. Personen mit Lernschwierigkeiten können nur die einfachen Items eines solchen Powertests beantworten. Das führt dazu, dass sie lediglich die ersten Items des Tests lösen können. Nach zwei bis drei gelösten Items sind die folgenden Items zu schwierig für diese Personengruppe. Der Test kann bzw. muss nun nach nur wenigen Items abgebrochen werden.

Wird ein Powertest von Personen mit einer hohen Kompetenz bearbeitet, beantworten auch diese zuerst eine Reihe leichter Items am Anfang. Anschließend folgen schwierigere Items, welche für diese Personengruppe in ihrer Schwierigkeit passender sind. Nach einem Testdurchgang haben jedoch auch diese Personen mit hohen Kompetenzen nur relativ wenige Fragen mit einer für sie passenden Schwierigkeit bearbeitet, da viele Items zu leicht für sie waren. Sowohl bei Personen mit niedrigen als auch mit hohen Kompetenzen können solche Powertests daher relativ gesehen nur wenige relevante Informationen liefern.

Eine Alternative hierzu bietet das adaptive Testen. Beim adaptiven Testen steht die Frage im Vordergrund, wie man mit möglichst wenig Items möglichst schnell die passenden Informationen bekommt. Das Ziel ist somit eine sehr effiziente Testung, bei der möglichst viel diagnostische Information über eine Testperson gesammelt werden kann (Frey, 2020). Die Hoffnung dabei ist, dass man auch Personen mit extremen Profilen (also sehr geringen oder hohen Kompetenzen) schneller und genauer messen kann. Natürlich können auch Personen mit mittlerem Kompetenzprofil einen adaptiven Test erfolgreich durchführen. Hier ist jedoch fraglich, inwiefern die Testung kürzer oder genauer ist, als es bei einer Powertestung der Fall wäre.

Adaptive Tests passen sich während der Testung an das Leistungsvermögen der getesteten Person an. Das Schwierigkeitsniveau wird dabei vom Antwortverhalten der getesteten Person beeinflusst: Während auf eine falsche Antwort ein leichteres Item folgt, wird nach einer richtigen Antwort ein schwierigeres Item angeboten. Auf diese Art und Weise pendelt sich das Schwierigkeitsniveau des Tests schnell auf dem höchstmöglichen Schwierigkeitsniveau ein, die die Testperson mit ihrem Leistungsvermögen noch bearbeiten kann (siehe Abbildung 14).

Abbildung 14
Darstellung eines Testdurchlaufs eines adaptiven Tests



Gestartet wird häufig mit einem leichten bis mittelschweren Item. Die Lösungswahrscheinlichkeit eines vorgelegten Items liegt durch die Auswahl der Items bei ungefähr 50 % und passt zum Leistungsvermögen der Person. Die Auswahl der zur Bearbeitung vorgelegten Items ist flexibel und im Gegensatz zu einem regulären Powertest nicht starr. Es gibt daher keine feste Reihenfolge, in der die Items bearbeitet werden müssen. Stattdessen bekommt z. B. jedes Kind einer Klasse eine unterschiedliche Reihenfolge an Items, sofern nicht alle Kinder exakt dasselbe Antwortverhalten zeigen.

Take Home Message:
Adaptive Tests passen die Auswahl der Aufgaben an das Leistungsvermögen der Person an. Dadurch benötigen adaptive Tests weniger Testzeit und messen genauer bei Personen mit sehr hohem oder sehr niedrigem Kompetenzen.

Adaptive Tests benötigen Items, deren Schwierigkeiten psychometrisch vorher mittels der Item-Response-Theory geprüft wurden. Für die Konstruktion eines adaptiven Tests muss jedes Item einem Itempool und einem Schwierigkeitswert zugeordnet werden, um später passend durch den Algorithmus gezogen werden zu können. Daher benötigen solche Tests eine empirische Prüfung mittels z. B. Schulklassen, um Ergebnisse für jedes Item zu erhalten (Magis et al., 2017). Die analoge Durchführung von adaptiven Tests ist zeitlich aufwendig und daher kaum in der Praxis anzutreffen (Kubinger, 2021). Daher hat sich schon früh eine digitale Variante, das computerisierte adaptive Testen (CAT), durchgesetzt. CAT ist eine Weiterentwicklung des adaptiven Testens, das sich auf die Testkonstruktion und Testdurchführung am Computer oder Tablet bezieht. Durch einen vorher definierten bzw. programmierten Algorithmus werden die jeweils nächsten Aufgaben für die Testperson immer so ausgewählt, dass sie die größte Aussagekraft über die Fähigkeiten der Person treffen können und dementsprechend am besten zu ihren Kompetenzen passen. Neben der Ziehung spezifiziert man anhand des Algorithmus unter anderem auch, wann die Testung beendet ist oder abgebrochen wird (Magis et al., 2017; Frey, 2020).

Aus pädagogischer Sicht sollte vor Beginn der Testung thematisiert werden, dass der Test sowohl sehr leichte als auch sehr schwere Aufgaben enthält und auch gute Schüler:innen nicht alle Items lösen können. Insbesondere für Kinder mit hohen Kompetenzen kann es ungewohnt sein, dass sie Aufgaben bekommen, welche sie zu 50% nicht lösen können. Für Kinder mit sonderpädagogischem Förderbedarf hingegen kann die Testsituation eher eine positive Erfahrung darstellen, da in den meist linear konstruierten Tests diese Kinder nur einen geringen Anteil an Aufgaben lösen würden. Adaptive Tests sind daher besonders für Kinder geeignet, welche nicht nahe dem Mittelwert der Norm liegen. Ebenso ermöglichen sie eine kürzere Testdurchführung bei vergleichbar hoher Reliabilität.

Hinweis:
Es gibt Lehrkräfte, die wünschen, dass ihre Schüler:innen immer alle Items lösen. Dann würde aber ein Deckeneffekt bestehen und die Kompetenz der Schüler:innen kann noch höher sein, als der Test misst. Ähnlich ist es beim Bodeneffekt, bei der eine Person kein oder kaum Items gelöst hat. Der Wunsch für jede Einzelmessung ist es daher, dass eine Person genug Items bearbeitet und hiervon sowohl Items löst als auch nicht löst. Erst dann kann das aktuelle Niveau eingeschätzt werden.

5.8 Lernspiele, Games und Übungen mit Rückmeldungen während des Tests

Bei Spielen und Lernübungen geht es auch um die Frage, was ist mein aktuelles (Lern-)Niveau und wie kann ich mich verbessern. Computerspiele sind in Levels angelegt, um Spieler:innen neue Herausforderung bei gleich hohem Interesse anzubieten. Für das Erlernen der Benutzung in Tutorials wird meist die Methode der direkten Instruktion (siehe Kapitel 7) gewählt und der/dem Spieler:in wird kleinschrittig und strukturiert die Benutzung erklärt. Viele Spiele enthalten insbesondere zu Beginn sehr viele einfache Aufgaben, um das Spielprinzip zu festigen und motivierend zu wirken. Ebenso wird Gamefication als zusätzlicher Verstärker (Fortschrittsbalken, Rangliste, usw.) verwendet, um die Motivation zu steigern.

Das erfolgreiche durchspielen von Computerspielen erfordert neben Ausdauer auch spezifische Kompetenzen. Als Serious Games werden Computerspiele bezeichnet, wenn sie neben Unterhaltung auch Lerninhalte vermitteln oder Kompetenzen trainieren. Hierzu

beinhalten solche Spiele Aufgaben, Lösungswege und Tipps. Je nachdem wie das Spiel angelegt ist, können die Daten solcher Spiele auch für die pädagogische Diagnostik herangezogen werden oder einzelne Teile einer solchen Spielwelt als diagnostische Messung betrachten.

Die Frage ist hierbei, ob die Messung im Spiel die Anforderung der psychometrischen Gütekriterien erfüllt und die Messungen mit Tests auf Papier nachvollziehbar, replizierbar und vergleichbar sind. So besteht dann die Erwartung, dass eine Person ähnliche Aufgaben wie im Computerspiel (bsp. Schachspielen, Rätsel lösen, schulische Aufgaben in Mathematik) auch in Tests auf Papier lösen kann.

Mehr Informationen:
Die LVD Goldmünzenjagd als Teil eines kommerziellen Lernspiels.

6 Leitfäden zur pädagogischen Diagnostik

Studierende des Lehramts fühlen sich vom Begriff Diagnostik sowie von Test- und Fördermaterial manchmal abgeschreckt und denken, dass diese Materialien zu „wissenschaftlich" für ein das praktische Handlungsfeld seien könnten. Ebenso wird von Studierende geäußert, dass die die Bewertung von Kindern oder die Festlegung von pädagogischen Entscheidungen ein Bauchgefühl von Expert:innen ist. Hierbei wird eine Art Meister:innenlehre sich vorgestellt, wobei man nur in der Praxis sich etwas abschauen kann. Genau für das Gegenteil steht jedoch dieses Buch für pädagogische Diagnostik. Pädagogische Entscheidungen sollten fairen, transparenten Regeln unterliegen und nicht dem reinen Bauchgefühl unterworfen werden. Wissenschaftliche Standards ermöglichen es allen Personen pädagogischer Diagnostik zu erlernen. Anhand von standardisierten Test- und Durchführungsbeschreibung kann jede Person einen Test durchführen.

Daher ist es empfohlen verschiedene Tests und Fördermaterialien in der pädagogischen Praxis auszuprobieren, um selbst ein/e geübte/r Praktiker:in zu werden.

Weitere Informationen zu Test- und Fördermaterial finden sie auch unter dem Youtube Kanal Medialab Sonderpädagogik Lernen

6.1 Wie wählt man einen Test aus?

Es gibt eine Vielzahl an Tests. Sucht man jedoch nach einem bestimmten latenten Konstrukt und für eine bestimmte Personengruppe, wird die Auswahl schnell kleiner. Für die Auswahl von Tests ist daher wichtig, dass die Fragestellung im diagnostischen Prozess vorher im Detail geklärt ist und auch die Anamnese und Beobachtungen schon in den diagnostischen Prozess eingeflossen sind. Ein Test ist dann passend, wenn dieser inhaltlich das beobachtbare Problem des Kindes erfasst und konkretisiert.
Das Kind muss dann im Messbereich des Tests liegen und auch die Anforderungen an den Test erfüllen. Hierbei ist auch zu fragen, wie umfänglich und genau das Konstrukt abgebildet werden muss. Soll die Intelligenz beispielsweise sprachfrei in vielen Dimensionen gemessen werden, oder genügt es eine Einschätzung der Intelligenz mit einem kulturfreien Gruppentest wie dem CFT (Siehe Anhang 11.1). Welche Durchführung ist im Rahmen der Schule möglich und welche Förderung ist anschließend möglich. Insbesondere in Förderdiagnostik als Prozessdiagnostik, sollte man Instrumente nutzen, welche man in einem größeren zeitlichen Abstand als Evaluation für schulische Kompetenzen verwenden kann (siehe 11.2 bis 11.4). Die folgenden Fragen können bei der Auswahl eines passenden diagnostischen Instruments helfen.

- Ausgangspunkt: Pädagogisches Problem
- Anamnese: Was ist das pädagogische Problem beim Kind und im schulischen Umfeld?
- Welche Hypothese haben Sie für das pädagogische Problem?
- Welche Instrumente und Beobachtungen erfassen Teile des pädagogischen Problems?
- Wie umfänglich muss der Tests das latente Konstrukt erfassen?
- Welche Vergleichswerte bietet das Instrument an und für welche Zielgruppe ist das Instrument entworfen?
- Bietet das Instrument Förderentscheidungen bzw. Empfehlungen und/oder gibt es verknüpftes Fördermaterial?

Wissen über den Test und seine Konstruktionsart wird benötigt, um die Testergebnisse zu verstehen und für die Gestaltung der Lernumgebung zu nutzen. Je nach Fragestellung und Bezug gibt es drei Bezugsnormen (Siehe Jungjohann & Gebhardt, 2022) nach welchen man Ergebnisse eines Tests vergleichen kann. So gibt es die kriteriale Bezugsnorm mit einer

objektiven Anforderung (kriterial), die individuelle Bezugsnorm, welche die Lernentwicklung einer Person betrachtet und die soziale Bezugsnorm, welche die Ergebnisse mit den Ergebnissen anderer Kindern oder einer Norm vergleicht. Je nach Bezugsnorm benötigt man andere Tests:

Kriteriale Tests:	Erreichen von Zielen
Schulleistungstests mit Normierung: (Sozial)	Feststellung einer Abweichung von der Norm
LVD:	Individuelle Lernentwicklung

Standardisierte Tests werden entworfen, um einen gewissen Lernbereich einer Zielgruppe zu messen. Wenn kein passender Test in einem individuellen Fall verfügbar ist, gibt es die Option, einen eigenen Test zu konstruieren. Dieser ist nicht empirisch geprüft, sollte aber dennoch den Ansprüchen einer fairen Probe genügen.

Weitere Informationen:
Neben der Testothek an der Universität gibt es auch weitere Datenbanken, welche Informationen zu verschiedenen Tests bereithalten, sowie die Homepages der einschlägigen Verlage:

Auflistung der Tests des Ifbq.hamburg
Open Test Archive
Open Access Repositorium für Messinstrumente des Gesis

6.2 Wie wird eine faire Probe oder einen informellen Test konstruiert?

Ein Test hat das Ziel zwischen den einzelnen Personen zu differenzieren. Daher sollten keine Personen entweder alle Items oder gar keine Items lösen können. Die Kompetenz dieser Personen kann man nicht einschätzen. So wäre es auch bei kriteriale Tests wünschenswert, dass zwar 90% des Tests gelöst wurden, aber auch einzelne Items falsch beantwortet wurden. Dies entspricht einem Wahrscheinlichkeitsorientierten Modell, in welchem auch Flüchtigkeitsfehler möglich sind.
Informelle Tests und Proben sollten den Ansprüchen der Gütekriterien genügen. Daher ist eine empirische Prüfung auch anhand der Daten einer Klasse oder besser anhand mehrerer Personen zu empfehlen. Die nachfolgenden Fragen im Leitfaden hängen zusammen und können auch in einer anderen Reihenfolge sinnvoll sein.

Was soll der Test messen?
Kompetenztests oder Skilltests messen möglichst genau ein latentes Konstrukt und verwenden dazu einen Inhaltsbereich, ein Aufgabenformat und dieselbe Instruktion, um **eindimensional** zu messen. Das ist aber insbesondere bei umfänglichen Fragestellungen nicht das Ziel einer Messung, da dann eine **multidimensionale Messung** das Ziel ist. So bestehen Wissenstests und Klausuren häufig aus mehreren Inhaltsbereichen und haben Aufgaben mit verschiedenen Instruktionen und Formaten. Wenn der Test aus mehreren Inhaltsbereichen und sehr unterschiedliche Aufgabenformate besteht, kann es sein, dass der Test entweder als multidimensional oder als Test bestehend aus mehreren Tests betrachtet werden muss. Korrelieren die verschiedenen Aufgabenbereiche eines Tests kaum miteinander, dann misst man unterschiedliche latente Konstrukte und sollte diese getrennt auswerten und betrachten. Je nach Fragestellung werden die latenten Konstrukte dann auch als Index verrechnet, wenn man das Vorgehen theoretisch begründen und sich der gemessenen Konstrukte bewusst ist. Daher ist zuerst zu fragen, ob ein latentes Konstrukt oder mehrere latente Konstrukte gemessen werden. Nimmt man für eine multidimensionale Testung lieber eine großen Test (Testbatterie) oder unterschiedliche Tests.

Welche Bearbeitungsgeschwindigkeit haben die Schüler:innen und wie viel Zeit besteht für die Probe bzw. den Test?
Die Zeit ist ein wichtiger Faktor. Meist steht in der Schule nur eine begrenzte Zeit zur Verfügung oder die Personen haben nur eine bestimmte Konzentrationsdauer für das optimale Bearbeiten der Aufgaben. Je nach Bearbeitungsgeschwindigkeit können Personen mehr oder weniger Items in einer gewissen Zeit bearbeiten. Daher gehört es zu den entscheidenden Rahmenbedingungen festzulegen, wie viel Zeit zur Verfügung steht und wie viele Items auch Personen mit Schwierigkeiten in dieser Zeit bearbeiten können sollen, um eine reliable Messung zu ermöglichen.
Ist die Bearbeitungszeit wie bei einem **Speedtest** Teil der Aufgabenstellung, da beispielsweise die Leseflüssigkeit gemessen wird?
Oder sollte die Zeit eine untergeordnete oder im optimalen Fall keinen Einfluss in einem **Powertest** haben, da die Personen, wenn dann an den einzelnen schwieriger werdenden Aufgaben scheitern? Meist wird eine Mischung gewählt, in dem es drei Gruppen von Schwierigkeiten pro Inhaltsbereich/Dimension als Powertest gibt sowie eine Zeitbegrenzung, welche eine maximale Bearbeitungszeit vorgibt. Diese Zeitbegrenzung beeinflusst im optimalen Fall nur Personen, welche sich der Teilnahme entziehen wollen. Ist die Zeitbegrenzung zu eng gewählt, beeinflusst sie den langsamsten Teil der Schüler:innen negativ.

Ist der Test eindimensional oder mehrdimensional?
Können die Aufgaben mit einem latenten Konstrukt gelöst werden oder benötigt man zur Lösung der Aufgaben weitere Kompetenzen? Benötigt man für die Lösung des Mathematiktests auch eine hohe Lesekompetenz? Wenn ja, sollte man diesen Einfluss kontrollieren und Items mit und ohne Leseanteil im Test separat bewerten.

- ➔ Empirische Prüfung mittels Korrelation der Items untereinander und in Bezug auf die Summenwerte der einzelnen Skalen.

Hat der Test ein Kriterium?
Welche Aufgabentypen sollten aufgrund des Lehrplans und dem bisher vermittelten Unterricht bereits mit wie viel Prozent gelöst werden? Welche Erwartungen bestehen für schwache Schüler:innen, für durchschnittliche Schüler:innen und für Schüler:innen, welche das Kriterium bereits erlernt haben? Auch kompetente Schüler:innen machen Flüchtigkeitsfehler, daher sollten die definierten Bereiche bzw. Rohwerte anhand der gelösten Aufgaben nicht zu eng definiert werden.

Wie wurden die Schwierigkeiten konstruiert und wie wirken diese im Test?
Wenn der Test als Speedtest konstruiert ist, sollten alle Aufgaben empirisch ähnlich schwer sein. Als Powertest sollten die Aufgaben aufsteigend schwieriger werden. Die angenommenen Schwierigkeiten der Items sind meist in den Ergebnissen insbesondere bei Neukonstruktionen nicht immer der theoretischen Annahme entsprechend. Dies kann entweder an einer falschen Einschätzung der Stichprobe liegen, denn auch erfahrene Lehrkräfte über- oder unterschätzen ihre Schüler:innen, oder an einzelnen Instruktionen oder Items, welche zusätzliche oder nicht erwartete Schwierigkeiten erzeugen. Dies liegt häufig an unbekannten Aufgabenformaten oder an zusätzlichen Distraktoren bei der Aufgabenstellung oder den Antwortkategorien bei Multiple Choice Aufgaben. Dies muss im Einzelfall geklärt werden. Items, welche nicht die gewünschte Kompetenz messen, darf man nicht mit den anderen Items verrechnen.

- ➔ Empirische Prüfung anhand der Itemschwierigkeit: Alle Personen, die das Item gelöst haben/alle Personen der Stichprobe

Wie führe ich den Test fair und objektiv durch?
War die Testdurchführung objektiv und hatten alle Personen ähnliche Chancen ihre optimale Leistung abzurufen? Wurden die Instruktionen von allen Personen verstanden?
Meist hilft es, mit mehreren Beispielitems, welche gemeinsam bearbeitet werden, ein Instruktionsverständnis sicherzustellen. Insbesondere im Bereich von Kindern mit sonderpädagogischem Förderbedarf kann man auch einen Test komplett zur Probe durchführen und erst eine später Parallelversion einsetzen und auswerten (Hessels & Schwab, 2015). Damit kann sichergestellt werden, dass die erste Durchführung nicht an den Instruktionen scheiterte.
Ebenso ist bei Kindern mit besonderen Bedürfnissen zu berücksichtigen, dass diese genug Zeit und Pausen hatten, um den Test mit ihrem optimalen Leistungsvermögen zu bearbeiten.

Wie verrechne ich den Test?
Im optimalen Fall sind die Rohwerte des Tests normalverteilt und die erreichten Rohwerte der Schüler:innen entsprechen den Anforderungen der vorher festgelegten Kriterien. Inwiefern man aus den empirischen Werten Schlüsse oder Noten zieht, muss immer auch theoretisch begründet sein.

Rohwerte von Tests können sie in Noten, „Bestanden"/ „Nicht Bestanden" oder andere Verteilungen transformieren. Meist stammen Empfehlungen und Richtwerte aus den heimlichen Lehrplänen der jeweiligen Einrichtung, da es häufig keine gesetzlichen Regelungen gibt. Häufig hört man von Lehrkräften, dass für den Erhalt der Note 4 mindestens 50% aller Aufgaben richtig gelösten sein müssen. Diese Aussage macht bei Tests mit Ratewahrscheinlichkeit (Multiple Choice Test) oder vielen leichten Aufgaben Sinn, da bei der Annahme einer Normalverteilung eines Tests und passende Aufgaben zum Lernniveau eher einer durchschnittlichen oder guten Note zu empfehlen ist. Die Note ist eine Bewertung und Rückmeldung für den Lernenden. Je nach Passung der Aufgaben zu den Lernenden und der Anzahl von leichten, mittleren und schwierigen Aufgaben können Tests unterschiedlich bewertet werden (Siehe Bezugsnormen).

Meine Empfehlung für einen pädagogischen Test zur Messung von erreichten Leistungen ist es, die Aufgaben anhand von kriterialen Richtlinien zu entwerfen und anhand von Aufgabenpools zu bewerten, ob der Lernbereich bestanden wurde oder nicht. Benötigt man eine soziale Bezugsnorm, um als Abschluss eines Lehrgangs eine Note zu begründen, dann setzt man den Test aus mehreren (leicht/mittel/schwer) oder gleichen Schwierigkeitsstufen zusammen und bewertet die Mitte mit einer durchschnittlichen Leistung.

Beispielaufgabe:

Als Trainer:in im Sportverein für Handball (Beliebiger Ballsport) möchten Sie die grundlegende Fitness zur Vorbereitung ihrer Jugendlichen (C Jugend) wissen, um die Trainingseinheiten zu planen.

Was soll der Test messen?

Im ersten Schritt müssen Sie festlegen, welche Form der Fitness gemessen werden soll. Welche Anforderung besteht im Handball und welche Übungen kennen Sie, um diese zu messen? Aus wie vielen verschiedenen Aufgaben sollte der Fitnesstest bestehen, um für die Trainingsvorbereitung genügend Informationen zu liefern? Welche Informationen wären ausreichend, um unnötige Übungen und weitere Tests zu vermeiden?

Ist der Test eindimensional oder mehrdimensional?

Ein eindimensionaler Test wäre eine größere Aufgabe oder mehrere kleinen Aufgaben, welche genau dieselbe Fertigkeit (Skill) testen. Dies könnte beispielsweise der Coopertest sein, bei dem die Sportler:innen 12 Minuten möglichst weit laufen können.

Dieser Test liegt zwischen der Mittel und Langstrecke und gibt Informationen zur Ausdauerleistung. Jedoch benötigt man Erfahrung, um das eigene Lauftempo richtig einzuschätzen. Da Handballspieler keine Leichtathleten sind, ist zu fragen, ob nicht handballspezifischere Aufgaben aussagekräftiger wären.

Welche Aufgaben könnten dies sein? ...

Verrechnung & Skalierung
Sobald die einzelnen Ergebnisse in Form von Rohdaten vorliegen, müssen Sie mehrere Tests verrechnen oder bei einem einzelnen Test bestimmen, wie sie die Ergebnisse interpretieren. Beim Coopertest gibt es eine Normierung, welche die Ergebnisse nach Altersgruppe in sehr gut bis sehr schlecht einschätzt. Ob diese Bewertung für ihre Gruppe von Handballspielern aussagekräftig ist, ist die Frage.

Für eine informellen Test zur Trainingsvorbereitung genügt jedoch meist eine grobe Einschätzung der Fitness über die Gruppe, da im weiteren Training die gemessenen Fähigkeiten im Handballspiel beobachtet werden kann und auch die Personen selbst einschätzen können, inwieweit sie selbst ihre Fitness verbessern wollen.

Für Trainer empfiehlt es sich daher dieselben Tests später nach der Vorbereitung mit einer fitten Handballgruppe zu absolvieren, um Vergleichswerte und eine bessere Interpretation der Testergebnisse für die nächste Vorbereitungszeit zu gewinnen.

6.3 Auswertung von standardisierten Screenings

Die Auswertung von standardisierten Tests ist anhand der Gütekriterien vorgegeben (Kapitel 4.8) und meist sehr genau im Manual beschrieben, um möglichst Objektive Ergebnisse zu bekommen. Nach der Durchführung eines Tests erhält die Lehrkraft Auskunft über die Anzahl richtig (Summenscore) und falsch gelöster Aufgaben (Fehlerscore). Werden die Anzahl der richtig gelösten Aufgaben den bearbeiteten Aufgaben gegenübergestellt, so ergibt sich die Lösungswahrscheinlichkeit des/der Schülers:in als prozentualer Wert. Die Anzahl der falsch gelösten Aufgaben in Relation aller bearbeiteten Aufgaben ergibt die prozentuale Fehlerwahrscheinlichkeit. Der Fehleranteil gibt also Auskunft darüber, wie viel Prozent der Aufgaben falsch beantwortet wurden. Dieser Wert gibt Auskunft darüber, ob die/der Schüler:in die Aufgaben langsam und korrekt oder sehr schnell mit vielen Fehlern löst. Ebenso gibt es Schüler:innen, welche in Klassentestungen viele Items überspringen. Daher ist die Anzahl der gelösten Aufgaben allein nicht immer der entscheidende Wert. Diese Rohwerte werden dann anhand einer Normwerttabelle auf Alters- oder Klassenstufen bezogene normierte Werte bezogen.

Standardisierte Screenings vergleichen dann das Ergebnis des Kindes mit einer Normgruppen. Im Gutachten findet man dann folgende Formulierung: „Die Ergebnisse der Schülerin im Subtest 1 sind in der Normwertetabelle für Grundschulen im mittleren Bereich zu verorten. Sie erreichte einen Rohwert von xx Punkten. Das entspricht dem T-Wert von xxx und dem T-Werte-Band von xx-xx und befindet sich unter/im/über dem Durchschnitt." Meist geht es bei solchen Tests darum festzustellen, ob sich das Kind in einer Risikogruppe befindet, was bei den meisten Tests unter Prozentrang 16 definiert wird.

6.4 Auswertung von Lernverlaufsdiagnostik

Die Lernverlaufsdiagnostik betrachtet den Lernprozess einer Person über mehrere Testzeitpunkte hinweg und erlaubt die Interpretation der Ergebnisse mithilfe der individuellen Bezugsnorm. Die persönliche Entwicklung wird als alleiniger Bezugsrahmen gesehen, um die Ergebnisse und Leistungen einzuschätzen. Grundlage ist also der individuelle Lernverlauf jedes/r einzelnen Lernenden ohne Blick auf den Lernverlauf von anderen Lernenden. Leitfrage für die Interpretation könnte demnach sein: „Wie hat sich der Lernende im Vergleich zum ersten Mal entwickelt?". So bezieht man sich auf die Veränderung der Leistung zu vorherigen Messzeitpunkten.

Die Art der Lernentwicklung kann bestimmt werden, wenn mehrere Messzeitpunkte (Empfehlung: mindestens 5 bei kurzem Abstand, 3-4 bei einem mehrwöchigen Abstand) vorliegen. Da man vorher ein Lernziel vor Augen hat, kann recht schnell bestimmt werden, ob ein Lernfortschritt, eine Lernstagnation oder gar ein Lernverlust erkennbar ist. Die graphische Darstellung des Lernverlaufs sollte vom Computer oder vom Testleiter in ein Koordinatensystem eingetragen werden. Meist erlaubt das Anlegen eines Lineals bereits, um die Lernverlaufskurve zu erkennen. Zusätzlich bieten manche Testverfahren wie Levumi die Möglichkeit einer qualitativen Auswertung. Qualitative Auswertungen sind individuell und beziehen sich auf die Aufgaben selbst. Anhand solcher Auswertungen können Sie die Fehler erkennen, welche Aufgaben oder welche Aufgabengruppen den Lernenden Schwierigkeiten bereiten.

Haben sie eine Prä-, Intervention und Postphase können sie den Trend eines Lernverlaufsgraphen mit der Methode „Tukey Tri Split" bestimmen. Der Tukey Tri Split wird gerade im Schulkontext aufgrund seiner leichten Handhabbarkeit häufig durchgeführt (Siehe Kapitel 5.6 und Kapitel 8).

7 Diagnostik und Förderung

Der Einsatz der pädagogischen Diagnostik hängt auch vom verwendeten didaktischen Modell ab. Aktuell sind die meisten Screenings und Schulleistungstests anhand einer sozialen Bezugsnorm entwickelt, um Schüler:innen zu identifizieren, die mehr Schwierigkeiten haben als andere Schüler:innen (also eine Bezugsgruppe). Diese Verfahren sind vom Unterricht unabhängig und messen besonders differenziert Kinder mit Leistungen im Messbereich von Prozentrang 50, also in der Norm. Trotzdem kann man diese Tests gut verwenden, um Kinder unter Prozentrang 16, also unterhalb des Normbereiches, zu identifizieren. Die Identifikation von solchen Risikokindern allein ist jedoch nur der erste Schritt. Für diese Kinder müssen weitere pädagogische Maßnahmen remedial oder im Unterricht getroffen werden.

Adaptive Förderung wird im Unterricht der allgemeinbildenden Schule häufig durch den Einsatz von verschiedenen Förderniveaus umgesetzt. Meistens werden drei Förderniveaus bzw. Fördergruppen durch die Lehrkraft unterschieden. Die Auswahl, Passung und auch Evaluation des Förderniveaus kann anschließend an die Einteilung während der Förderung selbst mittels Prozessdiagnostik bestimmt werden.

Eine enge Verzahnung von Diagnostik und Förderung wird aber erst bei individuellem Lernen ermöglicht. Dies wird erfüllt, wenn adaptive Förderniveaus individuell angepasst oder weiterentwickelt werden, um auch spezifische Bedürfnisse zu erfüllen. Möglich wird das durch den Einsatz von Standards und differenziertem Material (beispielsweise Lernleitern), welche mit der Diagnostik verknüpft sind. Für sonderpädagogische Förderung oder intensive Förderung als präventives Angebot benötigt man eine individuelle Förderung, welche sowohl aus individuellen Angeboten als auch aus gemeinsamen Angeboten besteht.

7.1 Direkte Instruktion

Direkte Instruktion ist ein wichtiger Teil der pädagogischen Arbeit und die Grundlage für das **remediale Arbeiten**. Mit einer Diagnostik wird ein konkretes Problem oder eine konkrete Schwierigkeit im Lernen aufgedeckt. Nun ist die Frage, mit welchen Methoden das Lernziel möglichst gut erreicht werden kann. Nach Grünke (2006) ist für Kinder mit Lernschwierigkeiten und einem sonderpädagogischen Förderbedarf die Methode der **direkten Instruktion** ein effektiver Lernweg. So kommt er in seiner Metaanalyse zu folgender Schlussfolgerung:

„Ein übungsbetontes, schrittweise aufeinander aufbauendes, feedbackbetontes, strategisch ausgerichtetes und lehrkraftzentriertes Vorgehen im Sinne einer Direkten und einer Strategieinstruktion ist der Schulreife, der Intelligenz und der Fähigkeit zur komplexen Verarbeitung sowie der Problemlösefähigkeit sehr zuträglich" (Grünke, 2006, S. 250).

Grünke (2006) zeigte, dass Interventionen zur Förderung der Wahrnehmung weniger effektiv sind als eine direkte Instruktion als Förderung. Stellt man somit spezifische Lernprobleme fest, ist es effektiv, diesen Bereich direkt zu fördern als stattdessen damit zusammenhängende Bereiche oder entfernte Lernbereiche. So hilft ein Motorik- oder Wahrnehmungstraining nur wenig im Bereich der schulischen Leistungen.

Insbesondere zum Erreichen von Fertigkeiten (Skills) oder bestimmten Verhaltensweisen sind Lernmethoden auf Basis des Behaviorismus und Kognitivismus eine effektive Ergänzung zum oder im regulären Unterricht. So gibt Hattie (2012) direkte Instruktion mit einer Effektstärke von Cohens d = 0.6 an. Hattie sieht hierbei besonders die Zusammenarbeit der Lehrkräfte bei der Planung der Instruktion und auch die Anleitung, Beobachtung und Rückmeldung zu den individuellen Lernwege der Schüler:innen zu den gesetzten Lehrzielen als wichtig an.

Direkte oder auch explizite genannte Instruktion ist eine stark lehrer:innenzentrierte Methode der Förderung, in welcher mehrere Instruktionen vorstrukturiert sind und direkt zum Lernziel führen. Direkte Instruktion ist kleinschrittig und direkt auf eine Strategie oder konkrete Fertigkeiten (Skills) gerichtet, welche verbessert werden sollen. Man kann diese auch im regulären Unterricht in verschiedenen Sozialformen wie dem Frontalunterricht und dem Unterricht in Kleingruppen bzw. Einzelunterricht sowie im offenen Unterricht einsetzen. Die Methode wird sowohl in Förderschulen als auch in der Inklusion verwendet. Hierbei wird die Instruktion direkt vermittelt und unmittelbar Feedback gegeben. Hierzu gibt es unter dem Begriff Direkte Instruktion unterschiedliche Konzepte. Direkte Instruktion wird meist auf die Forschung von Siegfried Engelmann bezogen, welcher mit Kollegen, das DISTAR Programm entwickelte (Becker & Engelmann, 1978).

Historisches Video:
Getting There: DISTAR Arithmetic System

Weitere Informationen:
Carnine, D. & Engelmann, S. (2016). Theory of Instruction

Bei der direkten Instruktion wird der Unterricht in kleinen Einheiten aufgebaut, welche sich systematisch aufeinander beziehen und sich explizit auf ein Lern- oder Verhaltensziel richten. Von der Lehrkraft wird das Lernziel bestimmt und das Material und die Form der Unterstützung individuell auf das Leistungsniveau der und das Lerntempo der Schüler:innen abgestimmt. Das traditionelle Ablaufschema für direkte Instruktion besteht aus drei Schritten (Fisher & Frey, 2013):

- I do it (Vorzeigen als positives Model)
- We do it (Gemeinsames Anwenden)
- You do it (Anwenden und Verstetigen)

Einführendes Video:
Vom National Center on Intensive Intervention zu direct Instruction in Mathematik (41:17)

Obwohl diese Methode sehr effektiv ist, wird sie im Unterricht eher selten oder nur in Ansätzen in einem Methodenmix eingesetzt. Insbesondere in Kleingruppen oder Einzelsitzungen wie beispielsweise in der Lerntherapie wird dieser Ansatz häufig angewendet.

Powell (2022) definiert im Podcast explizite Instruktion dagegen nicht mehr nur im engen Schema eines I do, We do und You do, in welchem das Rollenvorbild I do streng behavioristisch geprägt ist. Eher ist die Lernsituation interaktionistisch geprägt und das I do ebenfalls ein Austausch, welcher über die Modellierung der Lernumgebung zum We do wird. Über die Interaktion und das gerichtete und unmittelbare Feedback wird die Lernsituation mit maximaler Teilhabe und Erfolgsmöglichkeit gesteuert.

Beispiel Schleife binden:
Schleife binden ist eine zentrale Fertigkeit (Skill), welche alle Kinder in Deutschland benötigen, um beispielsweise Schuhe zu tragen. Aus diesem Grund steht es in der ersten Klasse sogar im bayrischen Lehrplan Plus im Fach Werken. Obwohl es im Lehrplan vorkommt, wird es jedoch in der Schule meist nur mit einem Arbeitsblatt abgedeckt. Manche Kinder können bereits im Kindergarten eine Schleife binden, während andere diese Fertigkeit erst im Laufe der Grundschule erwerben. Empirische Forschung fehlt zu dieser wichtigen Fertigkeit. Man kann aber annehmen, dass die wenigsten Kinder die Fertigkeit im Sinne eines entdeckenden Lernens erwerben. Eher wird dies über direkte Instruktion im Elternhaus, Kindergarten und Schule gelernt. Hierbei wird zuerst die Methode im Sinne des I do vorgezeigt und erklärt. Dies können die Eltern 1:1 oder per Video vorzeigen. Danach beginnt das We do, also die gemeinsame Übung (Guided practice) mit verschiedenen Schuhen oder

anderen und weiteren Hilfsmittel als Interaktion (Prompts (direkte Aufforderung) und Feedback). Danach wird die Begleitung stückweise zurückgefahren und die Hilfe immer weniger, bis das Kind erfolgreich auch selbst eine Schleife gebunden hat (Scaffolding). Wie bei jeder Fertigkeit benötigt das Kind danach auch weitere Unterstützung, um in allen Gegebenheiten auch unter Druck erfolgreich und schnell die Schleife seiner Schuhe zu binden (z. B. im Sport bei Schnürschuhen).

Nach Archer und Hughes (2011, S.4) wird explizite Instruktion in folgenden Möglichkeiten im Unterricht angewendet:

1. Review
    a. Review homework and relevant previous learning.
    b. Review prerequisite skills and knowledge.
2. Presentation
    a. State lesson goals.
    b. Present new material in small steps.
    c. Model procedures.
    d. Provide examples and non-examples.
    e. Use clear language.
    f. Avoid digressions.
3. Guided practice
    a. Require high frequency of responses.
    b. Ensure high rates of success.
    c. Provide timely feedback, clues, and prompts.
    d. Have students continue practice until they are fluent.
4. Corrections and feedback
    a. Reteach when necessary.
5. Independent practice
    a. Monitor initial practice attempts.
    b. Have students continue practice until skills are automatic.
6. Weekly and monthly reviews

Video zum Exkurs:
I Do, We Do, You Do: Scaffolding Reading Comprehension in Social Studies (6:16)
Take Home Message:
Direkte Instruktion ist eine kurze lehrkraftzentrierte Vermittlungseinheit. Man kann direkte Instruktion auch im Frontalunterricht einsetzen, aber Frontalunterricht ist nicht mit direkter Instruktion gleichzusetzen.

Als Beispiel einer direkten Instruktion auf Basis eines positiven Modells (Vorbild) und Unterstützungen durch direkte Aufforderung (Prompts) und Verstärker beschreibt Collins (2012, S.55) eine Intervention mit Time Delay:

„A trial of wait training would proceed as follows:
1. Get the attention of the learner.
2. Provide the reminder, "Wait if you do not know what to do, and I will help you."
3. State the task direction.
4. Wait a predetermined number of seconds (usually 3-5).
5. Deliver the controlling prompt.
6. Reinforce the learner if he or she makes a correct response, or restate the reminder to wait if the learner makes an incorrect response."

Direkte Instruktion vermeidet weitere oder zusätzliche kognitive Belastungen beim Lernen. Hintergrund ist nach Lebens und Lauth (2016), dass das Arbeitsgedächtnis bei Kindern mit Lernstörungen sonst kaum noch ausreichende Kapazitäten hat, um die Informationen im Langzeitgedächtnis zu speichern. Nach der Cognitive Load Theory (CLT) (Sweller, 1994)

erzeugt das Erlernen neuer Information eine kognitive Belastung, welche vom Vorwissen der Person, der Komplexität der Lernaufgabe und der Methode des Lernens abhängt. Direkte Instruktion baut auf behavioristischen Methoden und kognitivistischen Lerntheorien insbesondere der CLT auf.

Eine Übung auf Basis von direkter Instruktion kann nach Grosche et al. (2011) beispielsweise aus mittelschweren Aufgaben bestehen, welche erst durch schwere Aufgaben ersetzt werden sollen, wenn man 80% der Aufgaben gelöst hat. Macht ein Kind einen Fehler, korrigiert die Lehrkraft unmittelbar und zeigt die gewünschte Lösung, bevor das Kind über die Lösung nachdenkt. Richtige Lösungen des Kindes werden unmittelbar verstärkt.

Weitere Informationen:
Einführungsvideo zu Video zur Cognitive Load Theory (2:54)

Der Podcast Psychologie fürs Klassenzimmer von Dr. Benedikt Wisniewski zum Thema Cognitive Load Theory.

Nach Swanson (2001) sollte direkte Instruktion täglich eingesetzt werden, ist bei der Präsentation und der Übung lehrkraftzentriert und wird dabei formativ evaluiert. Hierbei unterscheidet Swanson (2001) in seiner Metanalyse zwischen direkter Instruktion, welche sich direkt auf eine isolierte Fertigkeit (Subskill) richtet, und strategischer Instruktion mit Strategievermittlung und Selbstmonitoring (Metakognition). Beide sind sehr effektiv für Kinder mit Lernschwierigkeiten.

Arbeitsauftrag:
Direkte Instruktion wird von manchen Praktiker:innen und Didaktiker:innen als Frontalunterricht bezeichnet. Erörtern sie diese Aussage.
Was ist direkte Instruktion und wie kann man diese im Unterricht einsetzen?
Wie hängt direkte Instruktion mit Diagnostik zusammen?

7.2 Direkte Instruktion und/oder Problem Based Learning

Auch wenn einige Vertreter:innen (Zhang, Kirschner, Cobern and Sweller, 2022) direkte Instruktion für die beste Form der Wissensvermittlung halten, gibt es genug Evidenz dafür, dass inquiry-based instruction bei konzeptionellem Wissen effektiv ist (de Jong et al., 2023). Selbstgesteuertes Lernen oder problemzentrierte Ansätze setzen eine vertiefte Beschäftigung mit den Inhalten voraus und helfen den Schüler:innen Inhalte zu vernetzen und in auch einen Transfer herzustellen.

Take Home Message:
Je nach dem angestrebten Ziel der Lehre und des Lernens, sowie Vorwissen und Kompetenzen der Lernenden sind instruktive oder konstruktive Lernformen oder eine Mischung beider Formen die beste Wahl.

Bei konstruktiven Methoden finden die Schüler:innen selbst verschiedene mögliche Lösungswege zur Bewältigung der Aufgaben. Diese Methode wird häufig in der didaktischen und fachdidaktischen Forschung gewählt. Anhand anregende Aufgaben finden Schüler:innen eigene Lösungswege, ergründen und reflektieren ihre Wege, um die dahinterliegenden Strukturen und Konzepte selbst zu erkennen. Beispielsweise nach dem Programm PIKAS des deutschen Zentrums für Lehrkräftebildung Mathematik dienen gute Aufgaben als Anregung „zur Auseinandersetzung mit problemhaltigen Fragestellungen sowie zur Erkundung grundlegender mathematischer Strukturen und Gesetzmäßigkeiten" (Deutsches Zentrum für Lehrkräftebildung, 2022). Die Intervention erfolgt in kleinen Gruppen, ist schülerorientiert und benötigt didaktisch gut ausgebildete Lehrkräfte. Bei solchen Verfahren geht es nicht um eine Fertigkeit (Skill), sondern um ein tieferes Verständnis der Mathematik und eine Kompetenzentwicklung.

Unterricht und Schule komplett nach direkter Instruktion oder weiteren Konzepten des des Behaviorismus wie beispielsweise im programmierten Unterricht aufzubauen (Klauer, 1964), ist nicht zielführend. Konstruktivistisches Lernen und Didaktik sowie problembasiertes Lernen hat sich berechtigterweise als ein vorherrschendes Konzept des Unterrichts durchgesetzt. Die verschiedenen Methoden haben unterschiedliche Ziele und ergänzen sich. So muss nicht jede Unterrichtseinheit möglichst offen und frei gestaltet sein. Denn Lernen und Unterricht sind komplex. Struktur, wie Classroom Management, spielt in der Durchführung von Unterricht eine entscheidende Rolle. Eine aktive und anregende Lernumgebung besteht daher sowohl aus instruktionsbasierten als auch aus konstruktivistischen Methoden. Zum Erreichen bestimmter grundlegender und skillbasierten Lern- und Verhaltensziele sind behavioristische Methoden insbesondere bei Schüler:innen mit sonderpädagogischem Förderbedarf effektiv und sollten individuell und passgenau verwendet werden. Instruktionsbasierte Methoden werden insbesondere als Einführung von Skills oder Strategien verwendet. Eine enge Verschränkung mit der Diagnostik ist daher unumgänglich.

Umso komplexer das zu erlernende Konstrukt ist, umso schwieriger ist ein Nachweis der Evidenz, insbesondere bei Kindern mit Lernschwierigkeiten oder auch mit sonderpädagogischem Förderbedarf. Ebenso zeigt sich in den Forschungsergebnissen, dass explizite Lernumgebungen mittels Instruktion stärkeren Leistungen, als implizite Lernumgebungen. So zeigten Grundschüler:innen in expliziten Lernumgebungen danach häufiger den Einsatz von Rechenstrategien als Schüler:innen aus Lernumgebungen des selbst entdeckenden Lernens (Heinze et al., 2020).

Literatur zur Vertiefung:

Powell, S., Hughes, E. M. & Peltier, C. (2022). *Myths That Undermine Maths Teaching.*

Hisgen, S., Barwasser, A., Nobel, K. und Grünke, M. (2022) Diagnostik. Konzeption und Evaluation praxisnaher Individualförderungen für Kinder und Jugendliche, die hinter den Ansprüchen des Bildungssystems zurückbleiben. Ein Praxisbeitrag. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 751-764). Universität Regensburg.

Weitere Informationen:

Archer & Hughes (2011). *Explicit instruction. effective and efficient teaching.* Guilford Press

7.3 Diagnostik und Förderung mit Levumi

Die Onlineplattform Levumi bietet rechnergestützt kurze, leicht zu handhabende Messverfahren an, sogenannte curriculumbasierte Tests, mittels derer Lehrkräfte Einblick in Lernverläufe der Schüler:innen in den Bereichen Lesen, Mathematik und Verhalten erhalten. Die Ergebnisse der Lernverlaufsdiagnostik bieten die Grundlage für pädagogische Entscheidungen für die Förderung und sind zudem Rückmeldungen zur Wirksamkeit des aktuellen Unterrichts und des pädagogischen Handelns. Die meisten Tests sind aktuell für den Bereich Grundschule und Sekundarstufe I entwickelt. Die Tests sind nicht in Klassenstufenform geordnet, sondern anhand von Kompetenzen und Niveaustufen, damit auch für Kinder mit Lernschwierigkeiten das passende Niveau ausgewählt werden kann. Die Plattform hält die Datenschutz Bestimmungen ein und es werden nur anonymisierte Datensätze zu wissenschaftlichen Fragestellungen ausgewertet. Informationen zu Levumi und zur Anwendung erfahren Sie unter:

- Levumi Handbuch

Ebenso gibt es Handbücher zu den einzelnen Bereichen:

- Lehrerhandbuch Deutsch
- Kurzhandreichung Mathematik
- Handbuch für Lehrkräfte im Anwendungsbereich Verhalten und Empfinden

Mehr Informationen finden Sie im Blog: https://www.levumi-blog.uni-kiel.de

- Video Erklärung zur Lernverlaufsdiagnostik
- Video Erste Schritte in Levumi
- Video Forschung und Idee hinter Levumi

Tests und Fragebögen

Einzelne Tests von Levumi.de sind als pdf veröffentlicht und unter Tests und Frageböge zu finden.

Einführende Literatur für die Praxis

- Vorstellung für die schulische Förderpraxis der Lernverlaufsdiagnostik in Deutsch und Mathematik mit der Onlineplattform www.levumi.de
- Schulische Förderpraxis mit und durch Lernverlaufsdiagnostik
- Lernverlaufsdiagnostik mit der Onlineplattform Levumi (2019)

Einführende Artikel

- Einführungsartikel in der Zeitschrift für Heilpädagogik (2016)
- Graphen der Lernverlaufsdiagnostik interpretieren und anwenden
- Vergleichender Artikel der Plattformen Lernlinie, Levumi und Quope (2022)

7.4 Ablaufschema Praktikum Förderdiagnostik

In Bayern gibt es wie in vielen Bundesländern zwei studienbegleitende Praktika an einem Vormittag in der Woche. Während das erste Praktikum zur Umsetzung der Förderdiagnostik dient, hat das zweite Praktikum die Aufgabe erste Unterrichtserfahrungen zu sammeln. Im Fokus des förderdiagnostischen Praktikums stand in meinem eigenen Studium in München (2003 bis 2008) das „DIE Modell" von Heimlich und damit verbunden ein umfassendes Einzelfallgutachten zu einem Kind mit Förderplan und der Umsetzung des Förderplans. Das umfassende Einzelfallgutachten ermöglichte es, eine genaue Statusdiagnostik eines Kindes auszuprobieren. Hierbei war aber schon die Auswahl des Kindes in der Förderschule schwierig, da das ausgewählte Kind möglichst den Anforderungen des Förderschwerpunkts Lernen entsprechen sollte, damit die verbundene Förderung in Mathematik, Deutsch oder dem Bereich der Metakognition bzw. Lernstrategien erfolgen konnte. Meist benötigte der/die Student:in viel Zeit für ein solches Einzelfallgutachten, denn dieses wurde benotet, und für die Durchführung und Evaluation der Förderung fehlte die Zeit.
Nachdem Förderung und Evaluation im Sinne einer Prozessdiagnostik mehr Informationen für die Diagnostik liefern und auch dem Kind mehr nutzen, haben wir den Schwerpunkt des studienbegleitenden Praktikums von der Status- zur Verlaufsdiagnostik umgestellt. Die umfassende Statusdiagnostik, wie man sie für eine Erkennung eines sonderpädagogischen Unterstützungsbedarfs benötigt, wird theoretisch anhand von Fallbeispielen besprochen.

Das Praktikum startet daher mit der grundsätzlichen Entscheidung, in welchem sonderpädagogischen Bereich eine klassen- bzw. kleingruppenorientierte Förderung erfolgen soll. In diesem Lernbereich wird die Klasse mittels eines Screenings getestet und die Leistungen und Arbeitszeugnisse der Kinder ausgewertet sowie der Unterricht beobachtet. Anhand der Ergebnisse des Screenings werden Kinder mit ähnlichen Förderprofilen ausgewählt (Probleme im Zehnerübergang, sehr langsame Leser...). Der **November** (siehe Tabelle 8) umfasst die Auswahl des Lernbereichs, das Screening der Kinder und eine Auswahl der Fördergruppe von 3 bis 5 Kindern. Im **Dezember** wird die aktuelle Lage der Kinder im Lernbereich mittels formativer Lernverlaufsdiagnostik dokumentiert und die Zone der nächsten Entwicklung festgestellt sowie die Förderplanung fixiert. Der **Januar** und der Beginn des **Februars** stellt die Interventionsphase dar, in der der aktuelle Lernbereich der Gruppe der Kinder im gemeinsamen Unterricht, in Kleingruppen und in Einzelförderung mit begründeten Methoden gefördert wird. Ende **Februar** findet dann mit der LVD eine kurze Evaluation des erreichten Lernzieles und ein Vergleich mit den vorgenommenen Zielen statt. Offene Fragen sind, welche weitere Entwicklung erwartet wird und welche Zugänge effektiv waren. Anhand der Dokumentation über die Diagnostik und Förderung in Form eines Portfolios erfolgt die Bewertung. Ziel der Lehre ist es dabei, dass die Dokumentation sowie das Vorgehen nachprüfbar und logisch begründet sind.

Tabelle 8
Ablaufplan des studienbegleitenden Praktikums Sonderpädagogik Lernen in Regensburg

Monate	Anwendung	Fragestellung
Nov.	Screening + LVD	Wo liegen die Probleme in Deutsch oder Mathematik?
Dez.	LVD + Beobachtung	Welche aktuelle Entwicklungsstufe hat das Kind?
Jan.	LVD + Förderung	Welche Förderung ist effektiv?
Feb.	LVD + Förderung	Wie sieht die weitere Entwicklung des Kindes aus?

8 Einzelfallforschung

In der Sozialforschung wird unter Einzelfallforschung die Analyse eines Einzelfalls verstanden. Dieser Fall kann entweder eine Einzelperson, Personengruppen, Organisationen/Netzwerke, ganze Gesellschaften bzw. Kulturen oder andere Formen sozialer Prozesse, Episoden, einzelne Situationen oder Ähnliches sein (Hering & Jungmann, 2014). Dieser Einzelfall wird anhand eines oder mehrerer Instrumente über die Zeit beobachtet und analysiert. Einzelfallforschung wird in einer Reihe von wissenschaftlichen Disziplinen verwendet, wie beispielsweise Psychologie, Pädagogik und Sonderpädagogik, in welcher wenige Personen für Untersuchungen zur Verfügung stehen.

In der sonderpädagogischen Forschung hat die Einzelfallforschung eine lange Tradition insbesondere im englischen Sprachraum (Grünke, 2012; Plavnick & Ferreri, 2013), da das Individuum in der sonderpädagogischen Diagnostik und Förderung im Vordergrund steht. Als Einzelfall muss für ein Kind mit sonderpädagogischem Unterstützungsbedarf die Frage geklärt werden, wie man den Lernerfolg oder die Teilhabe in der Schule sicherstellen kann bzw. welche Maßnahmen wirksam sein könnten und waren. Um die Frage der optimalen Förderung auch bei Personen mit Lernschwierigkeiten zu klären, benötigt man Studien und eine Klärung der Evidenz von Förderprogrammen. Welche Maßnahmen könnten für das betroffene Kind effektiv sein. Im Sinne des Konzeptes der Evidenz sollten dies Maßnahmen sein, welche theoretisch und empirisch ihre Wirksamkeit in mehreren Studien von unterschiedlichen Forscher:innengruppen nachgewiesen haben. Größere Untersuchungen von vielen betroffenen Personen mit spezifischen Schwierigkeiten sind aber meist schwierig zu finden. Umso spezifischer die Beeinträchtigung und der Kontext ist, umso weniger relevante Personen findet man für eine nötige Studie. Diese Form der Studien wird angewendet, wenn eine Studie aufgrund einer kleinen Population nur an wenigen Personen möglich ist, da es sich um eine spezielle Personengruppe oder partikuläre pädagogische Situationen handelt. Daher gibt es neben den Forschungsdesigns des Experiments als randomisierte Kontrollgruppenstudie (randomized controlled trial (RCT)) und Untersuchungen im Querschnitt oder Längsschnitt auch die kontrollierten Einzelanalysen.

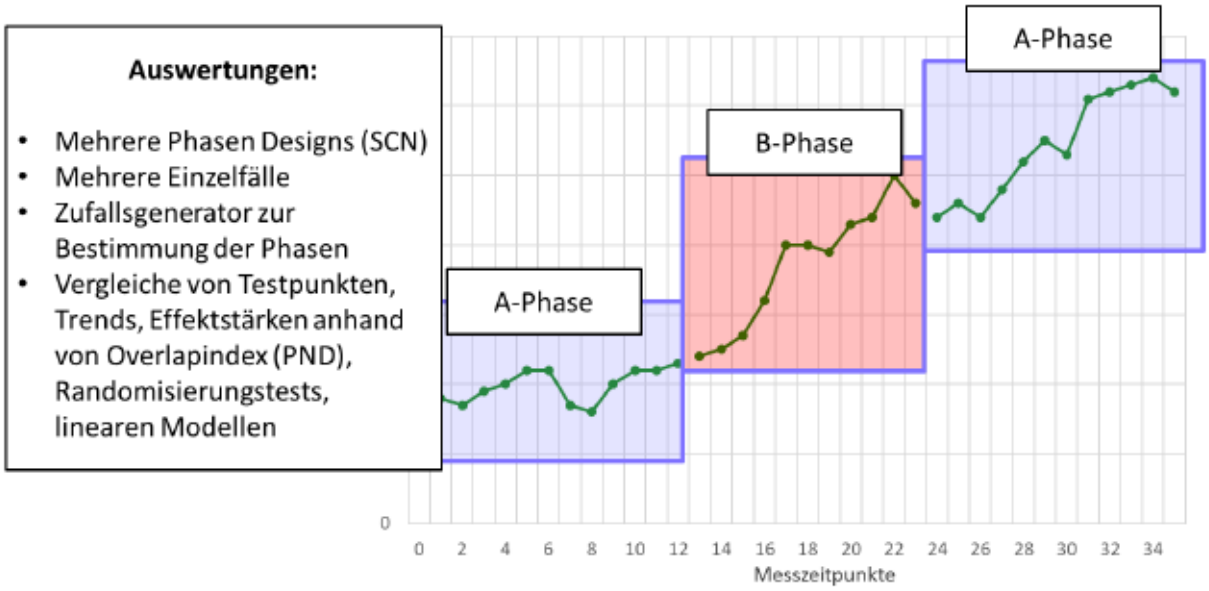
Take Home Message:
Einzelfallforschung analysiert einzelne Personen oder kleine Gruppen über eine längere Zeit anhand wissenschaftlicher Methoden.

Die Wirksamkeitsprüfung einer Intervention bei einem oder wenigen Kindern steht im Mittelpunkt. Diese wird standardisiert durchgeführt und anhand von mehreren Messzeitpunkten bei einem oder mehreren Einzelfällen ausgewertet (Grünke, 2012). Im single-case design (SCD) steht dabei der Vergleich einer bestimmten Intervention mit einer anderen Intervention oder keiner Intervention im Vordergrund (Jain & Spieß, 2012).

Einzelfallstudien können per Zufallsgenerator in zwei oder mehrere Phasen geteilt werden. Abbildung 15 zeigt beispielsweise ein ABA Design, welches neben der Kontrolle (A-Phase) auch eine Phase mit Intervention (B-Phase) enthält. In der A-Phase (Baseline-Phase, Grundratenphase) wird die abhängige Zielvariable (z. B. Leseflüssigkeit) unter Kontrollbedingungen (meistens: keine spezifische Intervention) mehrfach gemessen. Dabei ist das Ziel den aktuellen Verhaltens- oder Lernverlauf festzuhalten. In der B-Phase (Interventionsphase) wird die Experimentalbedingung eingeführt. Dies bedeutet, dass das Lese- oder Verhaltenstraining durchgeführt wird und mittels Verlaufsdiagnostik die Messungen der abhängigen Variable weiterhin festgehalten werden. Ist die Intervention erfolgreich, müsste eine deutliche Verbesserung sichtbar sein. In einer anschließenden weiteren A-Phase sollten die erworbenen Lern- bzw. Verhaltensfortschritte ohne Intervention weiter bestehen oder verbessert werden. Es besteht aber auch die Möglichkeit, dass die erworbenen Erfolge der B-Phase wieder vergessen wurden und es zu einem Leistungsabfall kommt. Mittels unterschiedlicher Verfahren (Non-Overlap-Indizes, Randomisierungstests,

Regressionsverfahren) können die Effektstärke und bei randomisierter Phaseneinteilung auch die Signifikanz geprüft werden.

Abbildung 15
Lernverlaufsgraph im Einzelfallforschungsdesign ABA



Phasendesigns haben die Annahme, dass Lernen und Verhalten möglichst wirkt und im beobachtbaren Zeitraum beobachtbar ist. Problematisch ist daher, wenn man den Beobachtungszeitraum zu kurz wählt. So kann es sein, dass sich der Lernerfolg verzögert und erst in der nächsten Phase zeigt. Dies muss beim Festsetzen der zeitlichen Abstände bedenken und in der Interpretation der Ergebnisse berücksichtigt.

Take Home Message:
Mehrere Phasen ermöglichen in der Einzelfallforschung Vergleiche innerhalb einer Person.

Weitere Informationen:

Das Werk von Julius et al. (2000) führt umfassend in die Einzelfallforschung unter dem Aspekt des Verhaltens ein.

Videoreihe von David Scheer zur Einzelfallforschung

9 Beratung und Elterngespräche

Beratung erfolgt zwischen einem Ratsuchenden und einem Berater. Der Ratsuchende sollte Vertrauen zu dem Berater haben und diesen auch frei wählen und wechseln können. Diese Voraussetzung ist in der Schule nur bedingt gegeben, da es meist ein beschränktes Beratungsangebot für spezifische Fragen von bestimmten Personen gibt. Im schulischen Kontext gibt es daher meist nur wenig freie Beratung. Angeboten werden meist beratende Gespräche, welche Problemstellungen und Herausforderungen klären.

Während Lehrkräfte in Beratungsgesprächen über die aktuelle Situation sprechen und dabei meist nur sich auf die letzten Wochen beziehen, haben Eltern einen langfristigen Blick und relativ wenig Informationen über den aktuellen Unterrichtsalltag. Eltern möchten daher verstehen, wie es ihrem Kind in der Schule geht und wie die individuelle Lernumgebung möglichst gut gestaltet werden kann. Daher haben Sie beispielsweise folgende Fragen:

- Wo steht mein Kind?
- Wie kann es sich verbessern?
- Was sind die konkreten Förderziele?
- Welche Maßnahmen trifft die Schule?

Pädagogische Diagnostik ermöglicht die Eltern umfassend zu informieren:

- Über den aktuellen Stand und die vergangene Entwicklung?
- Welche schulischen Maßnahmen stattgefunden haben und ob diese Erfolgreich waren?
- Welche Einschätzung zur zukünftigen Entwicklung die Lehrkraft hat?
- Welche Schritte die Schule und die Eltern sowie das Kind selbst anstoßen sollten?
- Welche Maßnahmen aus dem externen Unterstützungsfeld sinnvoll sind?
- Welche Förderziele festgelegt werden und wie diese erreicht werden?

Ziel von Elterngesprächen ist es, Probleme zu klären und Eltern einzubinden, um gemeinsam an einem Strang zu ziehen.

Mehr Informationen:
Beratung im Sonderpädagogischen Kontext von Andrea Schmidt

10 Ausblick

Im Alltag in deutschen Bildungsstätten wird Diagnostik oder die Verwendung von standardisierten und psychometrisch geprüften Instrumenten häufig entweder nur als Werkzeug für eine Klassifikation gesehen oder gar als lästige Dokumentationspflicht schon bekannter Probleme. Die eigentliche Idee der Diagnostik als Erkenntnisprozess und der damit verbundene Anlass zur Veränderung wird meist nicht erkannt oder gar ignoriert. Führt man Gespräche mit Praktiker:innen, dann bekommt man häufig die Antwort, dass ihnen die Notwendigkeit und die Bedeutung der pädagogischen Diagnostik bewusst ist, diese im Alltag aber aufgrund zahlreicher systembedingter oder auch persönlicher Gründe nur eine untergeordnete Rolle spielt. Aktuell sind die Erhebung und Auswertung von diagnostischen Daten für die Gestaltung von pädagogischen Maßnahmen im Alltag der Bildungseinrichtungen kaum präsent. Noch weniger wird eine gemeinsame Interpretation der Daten sowie ein damit verbundenes zielgerechtes weiteres Handeln auf Ebene der Personen oder gar der Institution durchgeführt. Im aktuellen Bildungssystem handeln bei Fragen zur Diagnostik einzelne Personen meist isoliert. Eine relevante und notwendige Vernetzung der an der Diagnostik Beteiligten, der Austausch von Daten und das Beschließen von gemeinsamen Lösungen findet nur unzureichend statt. Der Einsatz von Evaluation und Prozessdiagnostik auf Ebene von Teams oder eine Implementation auf Schulebene ist mir aus der Praxis nicht bekannt, aber in Bezug auf pädagogisches Handeln dringend notwendig.

Die pädagogische Praxis ändert sich langsam, aber stetig. Aufgrund des technischen Wandels wird es im digitalen Bereich immer einfacher, Daten zu speichern und diese systematisch auszuwerten. Daher gilt es weiter daran zu arbeiten, dass eines Tages im Alltag der Bildungseinrichtung der aktuelle (Lern-)Stand und die Entwicklung einfach erfasst werden können und ein darauf abgestimmtes pädagogisches Handeln dokumentiert und evaluiert wird. Im aktuellen analogen Alltag bearbeiten Kinder Aufgaben, ohne dass rechtzeitig eine Rückmeldung erfolgt. Meist erhalten Kinder eine Rückmeldung erst Wochen später, wenn im Alltag der Lehrkräfte Zeit für eine Korrektur möglich war. Insbesondere Kinder mit Lernschwierigkeiten benötigen jedoch viele kleine und schnelle Rückmeldungen, da ansonsten die Gefahr besteht, dass sie an den nicht passenden Herausforderungen scheitern und letztlich resignieren. Diese Problemlage wird sich zwar auch durch die digitale Technik und Bildung nicht vollkommen lösen lassen, aber zumindest sind automatische Auswertungen und Rückmeldungen leichter und einfacher im Alltag implementierbar. Ebenso sind im Digitalen die einzelnen Aufgaben und Lösungsversuche dokumentiert und digital bearbeitbar. Dadurch besteht nun ein einfacherer Zugang zu Daten und zur Analyse. Die bearbeiteten Aufgaben des Kindes können über eine längere Zeit betrachtet und ausgewertet werden. Ebenso kann direktes Feedback gestaltet werden und individuelle/ adaptive Lernumwelten entwickelt werden.

11 Anhang: Hinweise zu einzelnen Bereichen

Der Anhang des Buches ist als Verweis und Linksammlung für die Lehre gedacht und bildet daher fragmentarisch die einzelnen Bereiche ab. Die Kompetenzen, Leistungen und das Verhalten von Schüler:innen kann man mit unterschiedlichen Instrumenten erfassen. Jedes Instrument wurde für eine bestimmte Fragestellung und einen bestimmten Messbereich entwickelt und misst dementsprechend. Aus diesem Grund muss man sich als Anwender:in informieren und abschätzen, welches Instrument das gewünschte latente Konstrukt der betroffenen Person fair und ökonomisch misst, um die pädagogische Fragestellung zu beantworten. Ebenso sind nicht alle Produkte von Verlagen und anderen Anbieter:innen empfehlenswert.

Kostenpflichtige Testinstrumente kann man als pädagogische Einrichtung bei den Testverlagen wie beispielsweise Hogrefe oder Pearson erwerben. An Universitäten finden Sie daher neben der Bibliothek meist eine Testothek, um psychometrische Verfahren einsehen und ausprobieren zu können.

Mittlerweile veröffentlichen immer mehr Forscher:innen und Praktiker:innen ihre Instrumente auf Repositorien zur freien Nutzung wie beispielsweise dem Testarchiv. Solche Tests und Instrumente finden Sie am einfachsten durch eigene Open Access Suchmaschinen wie BASE. Ebenso veröffentlichen die einzelnen Forscher:innengruppen ihre Verfahren auf ihrer Webpage. Auch die Verfahren von Levumi.de finden sie auf unserer LMU Webseite. Die Verfahren des RIM Projektes unter Lernlinie. Nachfolgend werden daher insbesondere zusammenfassenden Testinformationen oder freie Verfahren verlinkt. Eine ausführliche Beschreibung der Verfahren findet sich auf den genannten Webseiten.

11.1 Kognitive Voraussetzungen und Intelligenz

Kognitive Fähigkeiten sind meist eine zentrale Fragestellung in der Diagnostik. Hierbei gibt es unterschiedliche latente Konstrukte, welche in den einzelnen Beiträgen des Herausgeberwerks vorgestellt werden:

Kuhl, J., Wittich, C. und Schulze, S. (2022). Intelligenz - Konstrukt und Diagnostik. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 131-145). Universität Regensburg.

Rauch, W. A. (2022). Exekutive Funktionen. In: *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Version 1.0.* Regensburger Beiträge zur Inklusions- und Sonderpädagogik, (4) (14). Regensburg, S. 163-173.

Schulze, S., Wittich, C. & Kuhl, J. (2022). Kognitive Lernvoraussetzungen - Aufmerksamkeit und Gedächtnis. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 147-161.). Universität Regensburg.

Die Diagnostik der Intelligenz wird vielfach für umfassende Einzelgutachten und häufig in der sonderpädagogischen Diagnostik zur generellen Klärung des sonderpädagogischen Unterstützungsbedarfs angewendet, während die weiteren kognitiven Fähigkeiten seltener einzeln explizit getestet werden. Häufig wird ein multidimensionales Einzelverfahren für Gutachten implizit oder explizit verlangt. Häufige Verfahren im Schulsystem sind der Nonverbalen Intelligenztest SON-R, Assessment-Battery for Children (KABC) oder der Wechsler Intelligence Scale. Ebenso ist der Culture Fair Test (CFT) möglich, welcher weniger

Dimensionen hat und als Gruppentests eingesetzt werden kann. Jeder dieser Tests baut auf ein anderes Intelligenzmodell auf und versteht sowie misst daher eine andere Ausprägung des latenten Konstruktes Intelligenz. Welches Verfahren für welche Personengruppe das optimale Verfahren ist, ist eine offene Diskussionsfrage der pädagogischen Praxis. Intelligenztests sind sehr umfangreich und messen daher über die Zeit sehr stabile Werte einer Person. Ebenso sind die Verfahren umfassend erprobt. Für die meisten Tests liegen vom Projekt Dia-Inform Testinformationen vor, welche nachfolgend verlinkt sind.

Kritisch anzumerken ist, dass Intelligenzmessung bei Gutachten sehr häufig angewendet wird und es fraglich ist, ob nicht ein umfassender und reliabler Schulleistungstests genügt. Eine weitere Schwierigkeit ist, dass in der Praxis leider häufig nur der IQ Wert für die Klassifikation betrachtet wird. Nur selten werden die einzelnen Ergebnisse der Dimensionen in der späteren Förderung berücksichtigt.

Beschreibungen von Verfahren zur Messung der Intelligenz

Renner, G. & Scholz, M. (2017). Testinformation zum Non-verbalen Intelligenztest SON-R 2½-7 (Dia-Inform Verfahrensinformationen 001-01).

Renner, G. & Schroeder, A. (2018). Testinformation zur Wechsler Intelligence Scale –Fifth Edition (WISC-V) (Dia-Inform Verfahrensinformationen 003-01).

Scholz, M., Renner, G. & Mickley, M. (2019). Testinformation zur Kaufman Assessment-Battery for Children – Second Edition (KABC-II) (Dia-Inform Verfahrensinformationen005-01).

Ergebnisse der Forschung

Intelligenztests werden eher selten in der pädagogischen Forschung eingesetzt, da sie als Einzeltests sehr aufwendig zu erheben sind. Meist wird der CFT als Gruppentest oder einzelne Skalen verwendet.

Heine et al. (2018). Psychometric properties of the CFT 1 R for students with special educational needs

Gebhardt et al. (2012). Heterogene Gruppen in mathematischen Grundfertigkeiten: Eine explorative Erkundung der Fähigkeiten im Grundrechnen in Integrationsklassen der 5. Schulstufe

Renner, G.; Irblich, D.; Schroeder, A. (2022). Factor Structure of the KABC-II at Ages 5 and 6: Is It Valid in a Clinical Sample? Children, 9, 645. https://doi.org/10.3390/children9050645

Interventionsforschung Intelligenz

Für das Training der Intelligenz selbst ist insbesondere das Denktraining von Klauer (1990) erprobt, welches jedoch in der Praxis selten angewendet wird.

Klauer (1990). Denktraining für Schulanfänger: Ein neuer Ansatz zur kognitiven Förderung

Hager et al. (1995). Induktives Denken und Intelligenztestleistung -Analysen zur Art der Wirkung zweier Denktrainings für Kinder

Exkurs Video:
Klaus Kubinger (30:18) Kriteriumsorientierte Diagnostik via Prototypen

11.2 Schriftspracherwerb & Lesen

Obwohl Lesen eine zentrale Kompetenz ist, wird aktuell immer noch in vielen Schulen zu wenig Unterstützung für Kinder mit Leseschwierigkeiten geboten. Daher ist die Leseentwicklung je nach Vorwissen und Vorläuferfähigkeiten sowie der Übungsanwendung sehr verschieden, was man an den Normen des ELFE II erkennt. Die Frage ist daher, wie man Leseschwierigkeiten möglichst frühzeitig erkennt und präventive Hilfe ermöglicht. Kostenpflichtige Standardverfahren sind in der Praxis das Inventar zur Erfassung der Lesekompetenz im 1. Schuljahr (IEL), der Salzburger Lese- und Rechtschreibtests (SLRT II) oder der Leseverständnistest für Erst- bis Siebtklässler (ELFE II).

Schmidt, B.M., Schabmann, Alfred & Hennes, A. (2022). *Leseverständnis diagnostizieren.* In: Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. *Version 1.0*. Regensburger Beiträge zur Inklusions- und Sonderpädagogik, (4) (38). Regensburg, S. 493-503.

Mayer, A. (2022). *Diagnostik bei Lese- und Rechtschreibschwierigkeiten.* In: Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. *Version 1.0.* Regensburger Beiträge zur Inklusions- und Sonderpädagogik, (4) (37). Regensburg, S. 479-492.

Ergebnisse der Forschung

Jungjohann et al. (2018). Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit

Jungjohann et al. (2021). Pilotierung von Leseflüssigkeits- und Leseverständnistests zur Entwicklung von Instrumenten der Lernverlaufsdiagnostik. Ergebnisse einer Längsschnittstudie in der 3ten und 4ten Jahrgangsstufe

Interventionsforschung Lesen

Kuhl, Euker, Ennemoser (2015). Förderung des lautorientierten Lesens bei Schüler:innen und Schülern mit intellektueller Beeinträchtigung

Ennemoser, Kuhl, Pepouna (2013). Evaluation des dialogischen Lesens zur Sprachförderung bei Kindern mit Migrationshintergrund

Beschreibungen von Verfahren zum Lesen

Renner, G. & Scholz, M., (2020). Testinformation ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler – Version II (Dia-Inform Verfahrensinformationen 006-01).

SiL-Levumi - Tests der Leseflüssigkeit zur Lernverlaufsdiagnostik - "Silben lesen" der Onlineplattform www.levumi.de

SinnL-N2-Levumi - Tests zum sinnkonstruierenden Satzlesen als Lernverlaufsdiagnostik - "Sinnkonstruierendes Satzlesen" der Onlineplattform www.levumi.de

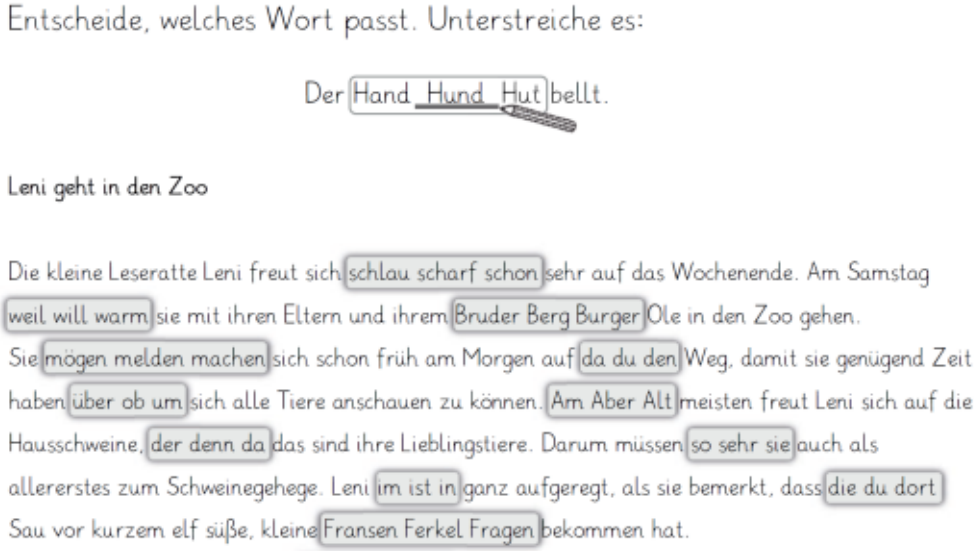
Beispiel-Verfahren zur Leseflüssigkeit:

Schnelles und korrektes Lesen von Silben, Wortlisten auf Zeit

Beispiel Maze Verfahren:

Ein beliebtes Verfahren zur curricular nahen Feststellung der Lesekompetenz sind MAZE Verfahren. Der Text enthält Lücken und mögliche Wörter. Der Tests enthält eine Speed- als auch je nach Schwierigkeiten der Lücke eine Powerkomponente. In Abbildung 16 ist ein Beispiel aus dem Leni Verfahren von Voß, Blumenthal, Höcker & Putnins (2019) auf Lernlinie.de dargestellt:

Abbildung 16
MAZE CBM Leni auf Textebene von Lernlinie.de


Entscheide, welches Wort passt. Unterstreiche es:

Der Hand Hund Hut bellt.

Leni geht in den Zoo

Die kleine Leseratte Leni freut sich schlau scharf schon sehr auf das Wochenende. Am Samstag weil will warm sie mit ihren Eltern und ihrem Bruder Berg Burger Ole in den Zoo gehen. Sie mögen melden machen sich schon früh am Morgen auf da du den Weg, damit sie genügend Zeit haben über ob um sich alle Tiere anschauen zu können. Am Aber Alt meisten freut Leni sich auf die Hausschweine, der denn da das sind ihre Lieblingstiere. Darum müssen so sehr sie auch als allererstes zum Schweinegehege. Leni im ist in ganz aufgeregt, als sie bemerkt, dass die du dort Sau vor kurzem elf süße, kleine Fransen Ferkel Fragen bekommen hat.

11.3 Rechtschreiben

Im Bereich Rechtschreibung gibt es eine große Auswahl an Statustests, Verlaufstests und Interventionen. Daher wurde bereits in einem Kapitel (siehe Kapitel 4.2) auf dieses Thema eingegangen.

Kostenpflichtige Standardverfahren sind in der Praxis die Hamburger Schreibprobe und die Oldenburger Fehleranalyse für die Klassen 1 und 2.

Ein freies Verfahren zur Messung der Rechtschreibung mittels Wort Diktaten liegt beispielsweise mit dem Rainer 1-4 für die Klassenstufen 1 bis 4 vor.

Weitere Informationen:
Hennes, A.-K., Schabmann, A., Vahlhaus-Aretz, D. & Schmidt, B.M. (2022) Schreibkompetenz diagnostizieren. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 505-516). Universität Regensburg,

Jungjohann, J. & Hüninghake, R. (2022). Förderplanung im Bereich Rechtschreibung. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung.. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 697-707). Universität Regensburg.

Ergebnisse der Forschung

Voß, Sikora & Mahlau (2017). Vorschlag zur Konzeption eines curriculumbasierten Messverfahrens zur Erfassung der Rechtschreibleistungen im Grundschulbereich. Empirische Sonderpädagogik.

11.4 Mathematik

Auch im Bereich der Mathematik sind Leistungsentwicklungen je nach Vorwissen und Vorläuferfähigkeiten sowie der Übungsanwendung sehr verschieden. In Mathematik kommt hinzu, dass einem Teil der Schüler:innen das mathematische Modellieren eher leicht fällt, während es für andere Kinder mit Lernschwierigkeiten eher eine Herausforderung ist. Ebenso spielt auch die Konzeption und Zugänglichkeit des aktuellen Mathematikunterrichts eine große Rolle, ob ein erfolgreiches Lernen und eine Kompetenzentwicklung stattfinden. Daher gibt es neben den standardisierten Mathematiktests wie der Demat Reihe, Basis-Math 4-8 oder der Marko Reihe viele informelle psychometrisch geprüfte oder ungeprüfte Verfahren.

Weitere Informationen:
Schulze, S., Wittich, C. und Vossen, A. (2022) Förderplanung im Bereich Mathematik. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 709-722). Universität Regensburg.

Ergebnisse der Forschung

Gebhardt, M., Ölkrug, K. & Tretter, T. (2013). Das mathematische Leistungsspektrum bei Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf in der Sekundarstufe. Ein explorativer Querschnitt der fünften bis neunten Klassenstufe in Münchner Förderschulen.

Fritz et al. (2018). Arithmetische Konzepte aus kognitiv-entwicklungspsychologischer Sicht.

Interventionsforschung Mathematik

Hecht, T., Sinner, D., Kuhl, J. & Ennemoser, M. (2011). Differenzielle Effekte eines Trainings der mathematischen Basiskompetenzen bei kognitiv schwachen Grundschülern und Schülern der Förderschule mit dem Schwerpunkt Lernen - Reanalyse zweier Studien.

Verfahren

Tests von der Arbeitsgruppe um Fritz & Ehlert

PsM100-Levumi N3. Test zu Platzhalteraufgaben zur Lernverlaufsdiagnostik-„Platzhalter Subtraktion (Zahlenraum 100)" der Onlineplattform www.levumi.de

DEMAT Reihe von Hogrefe

11.5 Verhalten und Erleben

Die Bewertung von Verhalten, Motivation und allen sogenannten non-kognitiven Outcomes gehört zu jeder pädagogischen Diagnostik. Entweder ist diese Diagnostik Teil der Abklärung oder Schwerpunkt der pädagogischen Fragestellung, wie beispielsweise beim Förderschwerpunkt emotionale und soziale Entwicklung (ESE). Daher finden sich zu dieser Fragestellung einige Artikel im Handbuch.

Weitere Informationen im Handbuch:
Roos, S. & Strumann, B. (2022). Diagnostik externalisierender Auffälligkeiten. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 369-382). Universität Regensburg.

Castello, A., Brodersen, G. und Grabowski, F. (2022) Förderdiagnostische Zugänge bei internalisierenden Auffälligkeiten. *In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 383-392). Universität Regensburg.

Zurbriggen, C. & Schmidt, P. (2022). Diagnostische Testverfahren im Bereich soziale und emotionale Entwicklung. Fokus Sozialverhalten, Sozialkompetenzen, Verhaltensstörungen. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 219-229). Universität Regensburg.

Schmidt, P. und Zurbriggen, C. (2022). Diagnostische Testverfahren im Bereich soziale und emotionale Entwicklung: Fokus Emotionen und Emotionsregulation. In M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 203-218). Universität Regensburg.

Ergebnisse der Forschung

De Vries et al. (2017). An assessment of measurement invariance in the 3-and 5-factor models of the Strengths and Difficulties Questionnaire: New insights from a longitudinal study.

Venetz et al. (2014). Entwicklung und erste Validierung einer Kurzversion des" Fragebogens zur Erfassung von Dimensionen der Integration von Schülern (FDI 4-6)" von Haeberlin, Moser, Bless und Klaghofer.

Huber & Rietz (2015). Direct Behavior Rating (DBR) als Methode zur Verhaltensverlaufsdiagnostik in der Schule: Ein systematisches Review von Methodenstudien.

Gebhardt et al. (2019). Measurement Invariance of a Direct Behavior Rating Multi Item Scale across Occasions.

Ausgewählte Verfahren

PIQ
Der PIQ ist das Standardinstrument zur Erfassung von Wohlbefinden, sozialer Partizipation und leistungsbezogenem Selbstkonzept in der Inklusionsforschung und wurde bereits in den ersten großen Studien zur Inklusionsforschung in der Schweiz von Urs Haeberlin und

Kolleg:innen (1991) entwickelt. Carmen Zurbriggen hat das Instrument gekürzt und weiterentwickelt. Es ist mittlerweile in viele Sprachen übersetzt, psychometrisch geprüft und in vielen Studien genutzt. Es gibt eine Version für Schüler:innen, Lehrer:innen und Eltern.

Weiter Informationen:
www.piqinfo.ch

Video:
Soziale Partizipation und Inklusion. Interview mit Carmen Zurbriggen (19:34)

SDQ
Der SDQ ist ein für die Praxis analoges kostenloses Einschätzungsinstrument, welches als Screening verwendet wird. Der SDQ wird in vielen Studien benutzt und ist daher eines der bestevaluierten Instrumente (Voß, & Gebhardt, 2017).

Weitere Informationen:
https://www.sdqinfo.org

DBR
Verlaufsdiagnostik für Verhalten in der Schule (DBR-PUTSIE)

12 Literatur

Archer, A. L., & Hughes, C. A. (2011). *Explicit instruction: Effective and efficient teaching*. The Guildford Press.

Autorengruppe Bildungsbericht (2014). *Bildung in Deutschland 2014. Ein indikatorengestützter Bericht mit einer Analyse zur Bildung von Menschen mit Behinderungen*. Bertelsmann. https://dx.doi.org/10.3278/6001820dw

AWMF (2018). S3-Leitlinie: Diagnostik und Behandlung der Rechenstörung. https://www.awmf.org/uploads/tx_szleitlinien/028-046l_S3_Rechenst%C3%B6rung-2018-03_1.pdf

Bender, F., Brandelik, K., Jeske, K., Lipka, M., Löffler, C., Mannhaupt, G., Naumann, C. L., Nolte, M., Ricken, G., Rosin, H., Scheerer-Neumann, G., von Aster, M. & von Orloff, M. (2017). Die integrative Lerntherapie. *Lernen und Lernstörungen 6*(2), 65–73. https://doi.org/10.1024/2235-0977/a000167

Berdelmann, K. (2016). „Sein Inneres kennen wir nicht, denn es ist uns verschlossen" – Schulische Beobachtung und Beurteilung von Kindern im 18. Jahrhundert. *Zeitschrift für Grundschulforschung, 9*(2), 9–23. https://doi.org/10.25656/01:15553

Becker, W.C. & Engelmann, Siegfried. (1978). Systems for basic instruction: Theory and applications. Handbook of Applied Behavior Analysis: Social and Instructional Processes. 325-377.

Betz, D. & Breuninger, H. (1998). *Teufelskreis Lernstörungen: Theoretische Grundlegung und Standardprogramm*. Beltz

Binet, A. (1912*). Neue Gedanken über das Schulkind*. Wunderlich.

Binet, A. & Simon, T. (1912*). A method of measuring the development of the intelligence of young children*. Courier.

Birkel, P. (2009). Rechtschreibleistung im Diktat - eine objektiv beurteilbare Leistung? *Didaktik Deutsch: Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur*, *14*(27), 5–32. https://doi.org/10.25656/01:21337

Blumenthal, Y., Kuhlmann, K. & Hartke, B. (2014). Diagnostik und Prävention von Lernschwierigkeiten im Aptitude Treatment Interaction-(ATI-) und Response to Intervention-(RTI-) Ansatz. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 73–94). Hogrefe.

Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, Massachusetts: Harvard University.

Brookhart, S. M. & McMillan, J. H. (2020). *Classroom assessment and educational measurement*. Taylor & Francis. https://doi.org/10.4324/9780429507533Brown-Chidsey, R. & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice*. Guilford Press.

Buchner, T. & Gebhardt, M. (2011). Zur schulischen Integration in Österreich - historische Entwicklung, Forschung und Status Quo. *Zeitschrift für Heilpädagogik*, *62*(8), 298–304. https://epub.uni-regensburg.de/43901/7/Buchner_%C3%96sterreich_1.pdf

Bühler, P. (2017). *„Diagnostik" und „praktische Behandlung". Die Entstehung der therapeutischen Funktion der* Schule. Beltz. https://irf.fhnw.ch/bitstream/handle/11654/25692/_buehler_2017_eeo.pdf

Bühler, P. & Hofmann, M. (2017). Education and Psychopathologization 1870-1940. *Bildungsgeschichte (IJHE), 7*(2), 133–141.

Bundschuh, K. & Winkler, C. (2019). *Einführung in die sonderpädagogische Diagnostik*. UTB.

Capovilla, C. & Kober, A. (2019). Intelligenzdiagnostik bei Kindern mit einer Sehbeeinträchtigung. *Empirische Sonderpädagogik*, *11*(1), 31–52. https://doi.org/10.25656/01:17769

Casale, G., Grosche, M., Volpe, R. J. & Hennemann, T. (2017). Zuverlässigkeit von Verhaltensverlaufsdiagnostik über Rater und Messzeitpunkte bei Schülern mit externalisierenden Verhaltensproblemen. *Empirische Sonderpädagogik*, *9*(2), 143–164. https://doi.org/10.25656/01:15010

Clemens, S. D. (1966). *Minimal brain dysfunction in children : terminology and identification ; phase one of the Three-Phase Project*. U.S. Department of Health, Education, and Welfare.

Collins, B. (2012). *Systematic Instruction for Students with Moderate and Severe Disabilities*. Brookes Publishing Co.

Cronbach, L. J. & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Deno, S. L. (1990). Individual differences and individual difference: The essential difference of special education. *The Journal of Special Education*, *24*(2), 160–173. https://doi.org/10.1177/002246699002400205

De Jong, T., et al. (2023). Let's talk evidence – The case for combining inquiry-based and direct instruction. *Educational Research Review. 39*. 100536. 10.1016/j.edurev.2023.100536.

De Vries, J. M., Rathmann, K. & Gebhardt, M. (2018). How does social behavior relate to both grades and achievement scores? *Frontiers in Psychology*, *9*, 857. https://doi.org/10.5283/epub.43566

Deutsches Zentrum für Lehrkräftebildung (2022) Verfügbar unter: https://pkas.dzlm.de/unterricht/gute-aufgaben [16.12.2022]

Ebenbeck, N., Jungjohann, J. & Gebhardt, M. (2022). *Testbeschreibung des Lesescreenings LES-IN für dritte inklusive Klassen.* Universität Regensburg. https://doi.org/10.5283/epub.53204

Ehlert A., Herzog M. & Fritz A. (2022). *Lost in Math – LoMa Testverfahren zur Erfassung arithmetischer Basiskompetenzen für die Sekundarstufe*. https://dyskalkulie-rechenschwaeche.de/tests/

Espin, C. A., Wayman, M. M., Deno, S. L., McMaster, K. L. & de Rooij, M. (2017). Data-based decision-making: Developing a method for capturing teachers' understanding of CBM graphs. *Learning Disabilities Research & Practice*, *32*(1), 8–21.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*(6), 359–374. http://dx.doi.org/10.1016/0001-6918(73)90003-6

Fisher, D. & Frey, N. (2013). *Better learning through structured teaching: A framework for the gradual release of responsibility*. Alexandria, VA: ASCD.

Fischer, R. A. (1935). *The Design Of Experiments*. Oliver & Boyd.

Frey, A. (2020). Computerisiertes adaptives Testen. In: Moosbrugger, H., Kelava, A. (Hrsg.) *Testtheorie und Fragebogenkonstruktion*. Springer. https://doi.org/10.1007/978-3-662-61532-4_20

Fuchs, L. S., Deno, S. L. & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision making. *Diagnostique*, *8*(3), 135–149. https://doi.org/10.1177/073724778300800301

Fuchs, D. & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, Why, and How Valid Is It? *Reading Research Quarterly, 41(*1), 93–99. http://www.jstor.org/stable/4151803

Garz, J. T. (2022). *Zwischen Anstalt und Schule. Eine Wissensgeschichte der Erziehung "schwachsinniger" Kinder in Berlin, 1845-1914*. Bielefeld: Transcript.

Gebhardt, M. (2022). *Inklusiv- und sonderpädagogische Pädagogik im Schwerpunkt Lernen. Eine Einführung.* Universität Regensburg. https://doi.org/10.5283/epub.45609

Gebhardt, M., DeVries, J.M., Jungjohann, J., Casale, G., Gegenfurtner, A. & Kuhn. T. J. (2019). Measurement Invariance of a Direct Behavior Rating Multi Item Scale across Occasions. *Social Sciences, 8*(2), 46. https://doi.org/10.3390/socsci8020046

Gebhardt, M., Heine, J. H., Zeuch, N. & Förster, N. (2015a). Lernverlaufsdiagnostik im Mathematikunterricht der zweiten Klasse. Raschanalysen zur Adaptation eines Testverfahrens für den Einsatz in inklusiven Klassen. *Empirische Sonderpädagogik*, (3), 206–222.

Gebhardt, M. & Jungjohann, J. (2020). Analyse der Lernausgangslage und der Lernentwicklung - Prozesse der Förderdiagnostik. In U. Heimlich & F. Wember (Hrsg.),

Didaktik des Unterrichts bei Lernschwierigkeiten: Ein Handbuch für Studium und Praxis (S. 367–380). Kohlhammer.

Gebhardt, M., Jungjohann, J. & Schurig, M. (2021). *Lernverlaufsdiagnostik im förderorientierten Unterricht. Eine Einführung in Anwendung, Testkonstruktion und Instrumente.* Reinhardt Verlag.

Gebhardt, M., Oelkrug, K. & Tretter, T. (2013). Das mathematische Leistungsspektrum bei Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf in der Sekundarstufe. Ein explorativer Querschnitt der fünften bis neunten Klassenstufe in Münch-ner Förderschulen. *Empirische Sonderpädagogik, 5*(2), 130-143.

Gebhardt, M., Sälzer, C., Mang, J., Müller, K. & Prenzel, M. (2015b). Performance of Students with Special Educational Needs in Germany. Findings from PISA 2012. *Journal of Cognitive Education and Psychology, 14*(3). 343–356. https://doi.org/10.1891/1945-8959.14.3.343

Gebhardt, M., Scheer, D. & Schurig, M. (2022). Sonderpädagogische Diagnostik. Eine Einführung. In. M. Gebhardt, Schurig, M. & Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 7–15). Universität Regensburg.

Genesee, F. & Hamayan, E. (1994). Classroom-based assessment. In F. Genesee (Hrsg.), *Educating second language children* (S. 212–239). Cambridge University Press.

Grosche, M., Hintz, A.-M. & Grünke, M. (2011). Direkte Instruktion in der Grundbildung. In K. Ratzke & A. Scholz (Hrsg.), *Alphabetisierung – Beratung –Chancen. Abschlussbericht zu einem Forschungs- und Entwicklungsprojekt* (S. 57–80). DIZ.

Grünke, M. (2006). Zur Effektivität von Fördermethoden bei Kindern und Jugendlichen mit Lernstörungen: Eine Synopse vorliegender Metaanalysen. *Kindheit und Entwicklung*, *15*(4), 239–254. http://dx.doi.org/10.1026/0942-5403.15.4.239

Grünke, M. (2012). Auswertung von Daten aus kontrollierten Einzelfallstudien mit Hilfe von Randomisierungstests. *Empirische Sonderpädagogik, 4*(3/4), 247–264. https://doi.org/10.25656/01:9301

Guthke, J. (1996). *Dynamisches Testen. Zur Psychodiagnostik der intraindividuellen Variabilität. Grundlagen, Verfahren und Anwendungsfelder.* Hogrefe.

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., Wayman, J. C., Pickens, C., Martin, E., & Steele, J. L. (2009). *Using Student Achievement Data to Support Instructional Decision Making.* United States Department of Education, Retrieved from https://repository.upenn.edu/gse_pubs/279

Hartung, N., Schurig, M., Vossen, A. & Gebhardt, M. (2021). Pädagogische Diagnostik im Rahmen des RTI-Modells. In J. Kuhl, A. Vossen, N. Hartung & C. Wittich (Hrsg.), *Evidenzbasierte Förderung bei Lernschwierigkeiten in der inklusiven Grundschule* (S. 28–39). Reinhardt Verlag.

Hartung, N., Vossen, A., Hecht, T. & Sinner, D. (2022). Diagnostik zur Feststellung des sonderpädagogischen Unterstützungsbedarf im Schwerpunkt Lernen. In. M. Gebhardt, Schurig, M. & Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 517-526). Universität Regensburg.

Hasbrouck, J. & Tindal, G. A. (2006). Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. *The Reading Teacher, 59*(7), 636–644. https://doi.org/10.1598/RT.59.7.3

Hattie, J. (2012). Visible learning for teachers: maximizing impact on learning. Routledge. https://doi.org/10.4324/9780203181522

Herrmann, S., Meissner, C., Nussbaumer, M. and Ditton, H. (2022), Matthew or compensatory effects? Factors that influence the math literacy of primary-school children in Germany. *Br J Educ Psychol, 92*, 518-534. https://doi.org/10.1111/bjep.12462

Heimlich, U. (2012). Inklusive Bildung für Menschen mit Behinderung - eine lebenslaufbegleitende Perspektive. https://www.bildungsbericht.de/de/schwerpunktthemen/pdfs/ expertenworkshop-2012-heimlich1112.pdf?mscl-kid=d521b061c7a811ec90a719afa16 deb99

Heimlich, U., Lutz, S., Willfert, K. (2021). *Ratgeber Förderdiagnostik: Feststellung des sonderpädagogischen Unterstützungsbedarfs im Schwerpunkt Lernen.* Persen Verlag.
Heimlich, U. (2022). Kind-Umfeld-Analyse und ökologische Modelle. In. M. Gebhardt, Schurig, M. & Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik* (S. 17–24). Universität Regensburg.
Heinze, A., Arend, J., Grüßing, M. & Lipowsky, F. (2020). Systematisch einführen oder selbst entdecken lassen? Eine experimentelle Studie zur Förderung der adaptiven Nutzung von Rechenstrategien bei Grundschulkindern. *Unterrichtswissenschaft, 48*, 11–34. https://doi.org/10.1007/s42010-019-00063-6
Hessels, M.G.P., Schwab, S. (2015). Kritische Aspekte bei Blickbewegungsanalysen am Beispiel der Evaluierung von Trainingseffekten im schlussfolgernden Denken bei Risikogruppen. *Empirische Sonderpädagogik 7*(3), 269-279.
Hillenschmidt, A. & Sander, S. (1993). *Kind-Umfeld-Diagnose. Ein ökosystemischer Ansatz*. Röhrig.
Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* (6. Aufl.). Beltz.
Jäger & Petermann (1999). *Psychologische Diagnostik. Ein Lehrbuch*. Beltz.
Jain, A. & Spieß, R. (2012). Versuchspläne der experimentellen Einzelfallforschung. *Empirische Sonderpädagogik*, *4*(3/4), 211–245. https://doi.org/10.25656/01:9300
Julius, H., Schlosser, R. W. & Goetze, H. (2000). *Kontrollierte Einzelfallstudien: eine Alternative für die sonderpädagogische und klinische Forschung*. Hogrefe.
Jungjohann, J., Diehl, K. & Gebhardt, M. (2019). *SiL-Levumi. Tests der Leseflüssigkeit zur Lernverlaufsdiagnostik - "Silben lesen" der Onlineplattform www.levumi.de.* ZPID (Leibniz Institute for Psychology) – Open Test Archive. https://doi.org/10.23668/psycharchives.4748
Jungjohann, J., & Gebhardt, M. (2023). Dimensions of Classroom-Based Assessments in Inclusive Education: A Teachers' Questionnaire for Instructional Decision-Making, Educational Assessments, Identification of Special Educational Needs, and Progress Monitoring. *International Journal of Special Education (IJSE), 38*(1), 131–144. https://doi.org/10.52291/ijse.2023.38.12
Jungjohann, J. & Gebhardt, M. (2021). Fragebogen zur Erfassung der diagnostischen Kompetenz von Lehrkräften in der inklusiven Schule (DaKI). Universität Regensburg. http://doi.org/10.5283/epub.49235
Jungjohann, J., Gebhardt, M. & Scheer, D. (2022). Understanding and improving teachers' graph literacy for data-based decision-making via video intervention. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.919152
Jungjohann, J., Gegenfurtner, A. & Gebhardt, M. (2018). Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit. *Empirische Sonderpädagogik, 10*(1), 100–118. https://doi.org/10.25656/01:15963
Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of Teacher Judgments of Preschoolers' Math Skills. *Journal of Psychoeducational Assessment, 30*(2), 148-159. https://doi.org/10.1177/0734282911412722
Karpicke, J. D. & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968. https://doi.org/10.1126/science.1152408
Klauer, K. J. (1964). *Programmierter Unterricht in Sonderschulen. Eine Einführung*. Marhold.
Klauer, K. J. (1975). *Lernbehindertenpädagogik*. Marhold.
Klauer, K. J. (1977). *Lernbehindertenpädagogik*. Marhold.
Klauer, K. J. (1978). Perspektiven der Pädagogischen Diagnostik. In K. J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik*. Pädagogischer Verlag Schwann.
Klauer, K. J. (1990). Denktraining für Schulanfänger. Ein neuer Ansatz zur kognitiven Förderung. *Praxis der Kinderpsychologie und Kinderpsychatrie, 39* (5), S. 150–156. https://doi.org/10.23668/psycharchives.11917

Klauer, K. J. (2005). W. Mutzeck & P. Jogschies (Hrsg.) (2004), Neue Entwicklungen in der Förderdiagnostik. Grundlagen und praktische Umsetzungen. *Zeitschrift für Pädagogische Psychologie*, *19*(3), 191–192. https://doi.org/10.1024/1010-0652.19.3.191.

Klauer, K. J. (2011). *Transfer des Lernens. Warum wir oft mehr lernen als gelehrt wird.* Kohlhammer.

Klauer, K. J. & Leutner, D. (2011). *Lehren und Lernen. Einführung in die Instruktionspsychologie*. Beltz.

KMK (2004): *Standards für die Lehrerbildung: Bildungswissenschaften.* https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf

KMK (2019). *Empfehlungen zur Digitalisierung in der Hochschullehre.* https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2019/BS_190314_Empfehlungen_Digitalisierung_Hochschullehre.pdf

KMK Statistik (2019). *Statistische Veröffentlichung der Kultusministerkonferenz. Dokumentation Nr. 223 – Februar 2020. Sonderpädagogische Förderung in Schulen 2009 bis 2018.* https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/Dok223_SoPae_2018.pdf

Kubinger, K. D. (2009). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Hogrefe.

Kubinger, K. D. (2021). Reducing the risk of lucky guessing as well as avoiding the contamination of speed and power in (paper-pencil) group-testing–illustrated by a new test-battery. *Psychological Test and Assessment Modeling, 63*(4), 458–468.

Lebens M. & Lauth, G. W. (2016). Direkte Instruktion. In G. W. Lauth & M. Grünke (Hrsg.), *Interventionen bei Lernstörungen: Förderung, Training und Therapie in der Praxis* (S. 418–428). Hogrefe.

Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Beltz.

Lord, F. I. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Lutz, S. (2023). Einzelfallraster für pädagogische Diagnostik (Version 0.1). Universität Regensburg. https://doi.org/10.5283/epub.53961

Lutz, S., Boschner, S. & Gebhardt, M. (2022). Data-based Decision Making (DBDM) in der inklusiven Diagnostik und Förderplanung. In. M. Gebhardt, M. Schurig & D. Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 33-42). Universität Regensburg. http://doi.org/10.5283/epub.53326

Lutz, S. & Gebhardt, M. (2021). *Fallbuch zum sonderpädagogischen Schwerpunkt Lernen*. Universität Regensburg. https://doi.org/10.5283/epub.46131

Lutz, S. & Gebhardt, M. (2023). Fallbuch zum sonderpädagogischen Schwerpunkt Lernen. Praxis-beispiele in Inklusion und Förderschule in Bayern. Universität Regensburg. https://doi.org/10.5283/epub.53980

Lutz, S., & Gebhardt, M. (2022). *Soziales und häusliches Umfeld von Schulkindern. Eine Einschätzung für Pädagog:innen (SHU-P).* Universität Regensburg. https://doi.org/10.5283/epub.47813

McCoy, A. (2011). *Teaching New Concepts: "I Do It, We Do It, You Do It" Method*. https://antoinemccoy.com/teaching-new-concepts/

Magis, D., & Barrada, J. R. (2017). Computerized adaptive testing with R: Recent updates of the package catR. *Journal of Statistical Software*, *76*, 1-19.

Mejeh, M. & Powell, J. (2018). Inklusive Bildung in der Schweiz - Zwischen globalen Normen und kantonalen Besonderheiten. *Bildung und Erziehung*, *71*(4), 412–431. http://dx.doi.org/10.13109/buer.2018.71.4.412

Moosbrugger, H. (2012). Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 227–274). Springer. https://doi.org/10.1007/978-3-642-20072-4_10

Moser Opitz, E. (2022). Diagnostisches und didaktisches Handeln verbinden: Entwicklung eines Prozessmodells auf der Grundlage von Erkenntnissen aus der pädagogischen

Diagnostik und der Förderdiagnostik. *Journal für Mathematik-Didaktik*, 1-26. https://doi.org/10.1007/s13138-022-00201-1

Moser Opitz, E., Pool Maag, S. & Labhart, D. (2019). Förderpläne: Instrument zur Förderung oder „bürokratisches Mittel"? Eine empirische Untersuchung zum Einsatz von Förderplänen. *Empirische Sonderpädagogik, 11*(3), 210–224. https://doi.org/10.25656/01:17780

Ogata, K. (2019). WISC-IV Factor Structures of Japanese Children with Borderline, or Deficient Intellectual Abilities: Testing Measurement Invariance Compared to Simulated Norm. *Psychology, 10*, 767–776. https://doi.org/10.4236/psych.2019.106050

Pawlik, K. (1982). *Diagnose der Diagnostik*. Stuttgart: Klett.

Plavnick, J. B. & Ferreri, S. J. (2013). Single-case experimental designs in educational research: A methodology for causal analyses in teaching and learning. *Educational Psychology Review*, *25*(4), 549–569.

Powell, S. (Moderierende). (2022). S2E4: Sarah Powell on Myths that Undermine Math Teaching. [Podcast-Folge]. In *Progressively incorrect*. https://educationrickshaw.com/2022/09/05/s2e4-sarah-powell-on-myths-that-undermine-math-teaching/

Prengel, A., & Liebers, K. (2005). ILeA 1. Individuelle Lernstandsanalysen. *Ein Leitfaden für die ersten sechs Schulwochen und darüber hinaus*. LISUM.

Prince, A. M., Yell, M. L. & Katsiyannis, A. (2018). Endrew F. v. Douglas County School District (2017): The US supreme court and special education. *Intervention in School and Clinic*, *53*(5), 321–324. http://dx.doi.org/10.1177/1053451217736867

Rasch, G. (1961) On General Laws and the Meaning of Measurement in Psychology. *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability, 4*, 321–333.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago.

Renner, G. & Scholz, M. (2017). *Testinformation zum Non-verbalen Intelligenztest SON-R 2½-7 (Dia-Inform Verfahrensinformationen 001-01)*. Pädagogische Hochschule Ludwigsburg.

Renner, G. & Scholz, M., (2020). *Testinformation ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler – Version II (Dia-Inform Verfahrensinformationen 006-01)*. Pädagogische Hochschule Ludwigsburg.

Renner, G. & Schroeder, A. (2018). *Testinformation zur Wechsler Intelligence Scale – Fifth Edition (WISC-V) (Dia-Inform Verfahrensinformationen 003-01)*. Pädagogische Hochschule Ludwigsburg.

Rost, J. (2004). *Lehrbuch Testtheorie -Testkonstruktion*. Bern: Huber.

Rost, J. & Spada, H. (1978). Probabilistische Testtheorie. In K. J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik* (S. 59–97). Pädagogischer Verlag Schwann.

Sälzer, C., Gebhardt, M., Müller, K. & Pauly, E. (2015). Der Prozess der Feststellung sonderpädagogischen Förderbedarfs in Deutschland. In P. Stanat, H. A. Pant, C. Gresch, M. Prenzel & P. Kuhl (Hrsg.), *Inklusion von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf in Schulleistungserhebungen. Forschungsstand, Herausforderungen und Perspektiven* (S. 129-153). Springer.

Sander, A. (1998). Kind-Umfeld-Analyse: Diagnostik bei Schülern und Schülerinnen mit besonderem Förderbedarf. In W. Mutzeck (Hrsg.), *Förderdiagnostik bei Lern- und Verhaltensstörungen. Konzepte und Methoden* (S. 6–24). Deutscher Studien Verlag.

Schneider, W., Lenhard, W. & Marx, P. (2019). Lern- und Verhaltensstörungen. In D. Urhahne, M. Dresel & F. Fischer (Hrsg.) Psychologie für den Lehrberuf (S. 565–585). Springer. https://doi.org/10.1007/978-3-662-55754-9_28

Scholz, M., Renner, G. & Mickley, M. (2019). *Testinformation zur Kaufman Assessment Battery for Children – Second Edition (KABC-II) (Dia-Inform Verfahrensinformationen005-01)*. Pädagogische Hochschule Ludwigsburg.

Schurig, M., Blumenthal, S. & Gebhardt, M. (2022). Continuous norming in learning progress monitoring—An example for a test in spelling from grade 2–4. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.943581
Schurig, M., Krause, S. & Gebhardt, M. (2020). *Verlaufsdiagnostik für Verhalten in der Schule (DBR-PUTSIE)*. http://dx.doi.org/10.13140/RG.2.2.33359.84641/2
Schwaighofer, M., Heene, M. & Bühner, M. (2019). Grundlagen und Kriterien der Diagnostik. In D. Urhahne, M. Dresel & F. Fischer (Hrsg.), *Psychologie für den Lehrberuf* (S. 471–491). Springer. https://doi.org/10.1007/978-3-662-55754-9_24
Schwanenberg (2009). *Leistungsentwicklung von Förder- und Regelschüler in der Sekundarstufe I.* Diplomarbeit. TU Dortmund
Sikora, S. & Voß, S. (2016). *Mathes 0. Schuleingangstest zur Erfassung arithmetischer Vorläuferfähigkeiten*. http://www.lernverlaufsdiagnostik.de
Strathmann, A., Klauer, K. J. & Greisbach, M. (2010). Lernverlaufdiagnostik - Dargestellt am Beispiel der Entwicklung der Rechtschreibkompetenz in der Grundschule. *Empirische Sonderpädagogik*, *2*(1), 64–77. https://doi.org/10.25656/01:9338
Swanson, H. L. (2001). Searching for the best model for instructing students with learning disabilities. *Focus on Exceptional Children, 34*(2), 1–15. http://dx.doi.org/10.17161/foec.v34i2.6785
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5
Tempel, T. & Pastötter, B. (2021). Abrufeffekte im Gedächtnis: Ein Überblick zur aktuellen Grundlagenforschung. *Psychologische Rundschau, 72*(4), 249–258. https://doi.org/10.1026/0033-3042/a000517
Tempel, T. & Sollich, S. (2023) Retrieval-based learning in special education. *Journal of Research in Special Educational Needs*, 00, 1– 7. https://doi.org/10.1111/1471-3802.12594
Tröster, H. (2019). *Diagnostik in schulischen Handlungsfeldern. Methoden, Konzepte, praktische Ansätze.* Kohlhammer.
United Nations (2006). *Convention on the Rights of Persons with Disabilities.* http://www.un.org/disabilities/default.asp?id=259.
Vaughn, S., Linan-Thompson, S. & Hickmann, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptionel Children, 69*(4), 391–409.
Vernooij, M. (2013). Sonderpädagogische Begutachtung. Thüringer Diagnostikkonzept zur Qualitätssicherung. Thüringer Ministerium für Bildung, Wissenschaft und Kultur. https://www.ngu.jena.de/wp-content/uploads/2015/09/Sonderpdagogische_Begutachtung_Vernooij.pdf
Voß, S., Blumenthal, Y., Mahlau, K., Marten, K., Diehl, K., Sikora, S. & Bodo, H. (2016). *Der Response-to-Intervention-Ansatz in der Praxis: Evaluationsergebnisse zum Rügener Inklusionsmodell*. Waxmann.
Voß, S. & Gebhardt, M. (2017). Verlaufsdiagnostik in der Schule. *Empirische Sonderpädagogik,* (2), 95–97.
Voß, S. & Gebhardt, M. (2017). Monitoring der sozial-emotionalen Situation von Grundschüler:innen und Grundschülern - Ist der SDQ ein geeignetes Verfahren*? Empirische Sonderpädagogik, 9*(1), 19–35. http://www.psychologie-aktuell.com/fileadmin/download/esp/1-2017_20170810/esp_1-2017_19-35.pdf
Walter, J. (2008). Adaptiver Unterricht erneut betrachtet. Über die Notwendigkeit systematischer formativer Evaluation von Lehr- und Lernprozessen und die daraus resultierende Diagnostik und Neudefinition von Lernstörungen nach dem RTI-Paradigma. *Zeitschrift für Heilpädagogik (6),* 202-215.
Weidlich, J. & Spannagel, C. (2014). Die Vorbereitungsphase im Flipped Classroom. Vorlesungsvideos versus Aufgaben. In K. Rummler (Hrsg.), *Lernräume gestalten – Bildungskontexte vielfältig denken* (S. 237–248). Waxmann.
Wolf, L. M. & Dietze, T. (2022). Ein Überblick über die Organisation der Feststellung von sonderpädagogischen Förderbedarfen in Deutschland. In. M. Gebhardt, Schurig, M. & Scheer (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik: Grundlagen und*

Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung. Regensburger Beiträge zur Inklusions- und Sonderpädagogik (S. 325–344). Universität Regensburg.

World Health Organization. (1980). *International classification of impairments, disabilities, and handicaps: a manual of classification relating to the consequences of disease, published in accordance with resolution WHA29.35 of the Twenty-ninth World Health Assembly, May 1976*. World Health Organization. https://apps.who.int/iris/handle/10665/41003

World Health Organization. (2001). *International classification of functioning, disability and health: ICF*. World Health Organization. https://apps.who.int/iris/handle/10665/42407

Wilbert, J. (2014). Vermittlung von Basiskompetenzen zum Rechnen. In G. Lauth, M. Grünke, & J. Brunstein (Hrsg.), *Interventionen bei Lernstörungen: Förderung, Training und Therapie in der Praxis* (S. 209–219). Hogrefe.

Zhang, L., Kirschner, P., Cobern, W. & Sweller, J. (2022). There is an Evidence Crisis in Science Educational Policy. *Educational Psychology Review. 34*. 10.1007/s10648-021-09646-1.