

Verbessern Leistungstests die Schulqualität?

Die Antwort gleich zu Beginn: Im engen Sinne – NEIN. Doch letztlich können Sie zur Optimierung des Nutzens von Leistungstests beitragen.

In der Zwischenzeit hat sich herumgesprochen, dass Check 5 kein Gesundheits-Check für alternde Lehrpersonen ist, Klassencockpit nicht die Pilotenkanzel des schuleigenen Flugzeugs meint, Stellwerk der SBB nicht zur Steuerung des Zugverkehrs dient und PISA nichts mit dem schönen italienischen Städtchen zu tun hat, sondern, dass es sich um Schulleistungstests mit unterschiedlichen Funktionen handelt.

Die Begriffe tauchen oft mit Versprechen im Zusammenhang mit der Qualität von Schule, Unterricht oder Bildungssystem auf. Die Ergebnisse von internationalen Schulleistungstests wie PISA oder TIMSS (Mathematikstudie) haben die Öffentlichkeit alarmiert und zu fetten News in den Medien geführt: Die Schülerinnen und Schüler in der Schweiz sind dumm («Blick» 5.12.2001). Vor allem Schülerinnen und Schüler mit Migrationshintergrund oder sozial benachteiligte Kinder schneiden in den Leistungstests schlecht ab. Der Aufschrei aus verschiedenen Kehlen war gross: Volkswirtschaftlich wird die Konkurrenzfähigkeit als gefährdet betrachtet, Politikerinnen und Politiker sehen die Teilnahme an der Demokratie gefährdet, Elternvereinigungen nutzen die Gunst der Stunde und fordern die freie Schuwahl als Mittel gegen den Schlendrian der Volksschule bzw. zur Steigerung der Schulqualität, aus finanzieller Sicht wird das Kosten-Nutzen-Verhältnis beklagt und schliesslich fühlen sich die Lehrpersonen an den Pranger gestellt und in ihrer Arbeit einmal mehr nicht wertgeschätzt. Diese unterschiedlichen Anspruchsgruppen haben eines gemeinsam: Den Ruf nach mehr Qualität in der Schule ... oft verbunden mit einer ideologisch gefärbten Zusatzbotschaft. Doch mit welchem Allheilmittel soll dieses Ziel erreicht werden?

Vom Wechsel der Richtung beim Denken: Von der Input- zur Output-Steuerung im Bildungswesen

Die Öffentlichkeit mit ihren verschiedenen Anspruchsgruppen hat sich in der Vergangenheit mit Ausnahme der Rekrutenprüfung kaum für die Wirkung, den Output, des Bildungssystems interessiert. Gesteuert wurde das Bildungssystem nach dem Prinzip «Umso mehr, umso besser», sprich mehr Ressourcen im Sinne von besseren Lehrplänen, neuen Lehrmitteln, kleineren Klassen, ausgebauter Infrastruktur, mehr Lektionen, besser qualifizierten Lehrpersonen, regelmässigen Inspektoraten, differenzierten Gesetzessammlungen. Die Qualität der Schule sollte durch den üppigen Ausbau des Inputs gesteigert werden. Und die Schweiz tat viel zur Steigerung des Inputs: 26 filigrane kantonale Lehrpläne, im Vergleich zu den Nachbarstaaten hohe Stundenverpflichtungen für die Schülerinnen und Schüler wie auch für die Lehrpersonen, kantonale

Lehrmittelverlage zur Herstellung obligatorischer Lehrmittel, Schulhausbauten mit Designerstühlen im Lehrerzimmer und High-End-ICT-Ausstattung. Eine OECD-Studie (2006) zeigt, dass die Schweiz über eines der teuersten Bildungssysteme verfügt. Der Glaube an dieses Input-Prinzip war so tief verwurzelt, dass es die Schweiz bis 1995 zu TIMSS (Third International Mathematics and Science Study) nicht für nötig befand, an einer internationalen Vergleichsstudie teilzunehmen. Auch im Bildungsbereich glaubte die Schweiz an ihre Rolle als Sonderfall. Auf dieser Prämisse sind der PISA-Schock und der Paradigmawechsel von der Input- zur Output-Steuerung zu verstehen. Das Bildungssystem mit seinen Schulen wird an seiner Effizienz und Effektivität gemessen. Nicht mehr über Input-Variablen soll das Bildungssystem gesteuert werden, sondern über den Output in der Form von Schulleistungsergebnissen und den Outcome in Form von längerfristigen Wirkungen wie Studien- und Berufserfolg.



Der Wechsel des Denkens ist erkennbar an folgenden Merkmalen:

- Dezentralisierung: Neuaufteilung der Verantwortung zwischen Bund, Kanton, Schulträger und Einzelschule
- Bildungsmonitoring: Steuerung durch Zielvorgaben in Form von Leistungsvereinbarungen und Bildungsstandards und der Überwachung der Wirkung durch Bildungscontrolling und Testsysteme.
- Teilautonome Schule: Erhöhung der Eigenständigkeit der Einzelschule durch Übertragung der pädagogischen, organisatorischen und teilweise personellen Verantwortung unter der Führung einer Schulleitung.
- Qualitätsmanagement: Einführung von Qualitätsmanagementansätzen in der Schule inkl. Selbst- und Fremdevaluation.
- Professionalisierung: Verlagerung der Aus- und Weiterbildung der Lehrpersonen auf die Tertiärstufe (Pädagogische Hochschulen), Aus- und Weiterbildung von Schulleitungen und Berufsstandards für die Schulaufsicht.

Mit diesem Wechsel des Denkens gewinnen standardisierte Schulleistungstests eine zentrale Bedeutung.

Was ist ein schulischer Leistungstest und was misst er?

Im Bildungsbereich bedeutet «Test» ein wissenschaftliches Verfahren zur Untersuchung eines oder mehrerer Persönlichkeitsmerkmale mit dem Ziel, über dieses Persönlichkeitsmerkmal eine quantitative Aussage zu machen. Ein Test muss verschiedene Gütekriterien der Testtheorie erfüllen:

- Objektivität: Das Testergebnis muss unabhängig von der untersuchenden Person und Situation sein. Verschiedene Urteiler kommen zu demselben Ergebnis.
- Zuverlässigkeit (Reliabilität): Zuverlässig ist ein Test dann, wenn sich bei wiederholter

Messung das Testergebnis nicht verändert. Die wiederholte Beurteilung derselben Schülerleistung durch dieselbe oder andere Beurteiler führt zu demselben Ergebnis.

- Gültigkeit (Validität): Validität bedeutet, dass der Test das Merkmal einer Person misst, das er vorgibt, zu messen. Je nach der Art des Merkmals kann dessen Messung unterschiedlich schwierig sein. So ist die Körpergrösse einfacher zu messen als Intelligenz. Die Konstruktion valider Tests ist eine hoch professionelle und letztlich teure Angelegenheit.

Standardisierte schulische Leistungstests umfassen von Expertinnen und Experten ausgewählte und entwickelte Testaufgaben, die sich an Lernzielen orientieren. Doch was messen diese Testaufgaben? Messbar ist nicht das Lernen, sondern letztlich nur eine sichtbare Leistung in einer bestimmten Situation zu einem bestimmten Zeitpunkt als Ausschnitt des gesamten Wissens. Aufgrund der gezeigten Leistung wird auf das erlernte Wissen und Können einer Schülerin oder eines Schülers geschlossen. Zur Bewertung wird die individuelle Leistung mit einer repräsentativen Vergleichsgruppe oder mit einem Lernziel verglichen und durch eine Zuordnung bewertet.

An dieser Stelle soll nur auf einige grundlegende, problematische Punkte hingewiesen werden:

- Schulische Leistungstests sind Situationsaufnahmen, und die Bewertung kommt durch Quervergleiche zustande. Sie geben keine eindeutigen Hinweise über die zukünftige Entwicklung einer Leistung und deren Förderung.
- Die gezeigte Leistung ist abhängig von der Art der Aufgabenstellung (z. B. Sprache) und der Situation wie mündliche Prüfung oder Zeitdruck. Als Ergebnis von Lernprozessen sind Testleistungen immer Momentaufnahmen und keine Grundlage für Typologierungen.

- Es besteht die Gefahr, die Leistungsbereiche zu messen, die einfach messbar sind. Damit erhalten diese eine hohe Bedeutung und verdrängen andere Bereiche wie soziale Entwicklung, musische Bereiche oder Problemlösefähigkeit.
- Die Interpretation der Testergebnisse durch die Lehrperson und die anschliessenden Folgerungen für den Unterricht insgesamt und die einzelnen Schülerinnen und Schüler ist höchst anspruchsvoll. Tests nur zur Selbstvergewisserung und Neugierde sind nur eine Belastung für die Lernenden und daher zu unterlassen.

Wozu welche Tests? – Funktionen und Bereiche von Schulleistungstests

Die Klärung der Funktion eines Tests ist zentral und muss zu Beginn eindeutig geklärt werden. Vereinfacht lassen sich folgende Funktionen und Bereiche von Leistungstests unterscheiden (s. Abbildung auf Seite 6).

Hinweis: Die angegebenen Schulleistungstests dienen als Beispiele zur Illustration und sind nicht wertend gemeint.

In der Fülle der Schulleistungstests die Übersicht zu behalten, ist schwierig. Zudem ist eine eindeutige Zuordnung der Funktionen oft nur schwer möglich. Dies verdeutlicht gleichzeitig ein Problem: Die Überlagerung verschiedener Funktionen in einem Test führt dazu, dass die «mächtigere» Funktion dominiert und die beispielsweise fördernde Funktion unterlaufen wird. Dies unterstützt nicht nur die Haltung «Teaching to the test», sondern die Lehrpersonen wenden mehr Zeit für die getesteten Fächer auf. Zusätzlich zu den Zeitverschiebungen zwischen den Fächern verändert sich auch der Lehr-Lernprozess. Da Tests oft einfache Wissensfragen und wenige komplexe Problemaufgaben umfassen, kon-

Bereiche	Funktionen		
	Rechenschaft/ Kontrolle	Qualifikation/ Selektion	Unterrichtsentwicklung/ Förderung
Schulsystem (national, kantonal)	PISA, Stellwerk, Jahresabschlussprüfung	Eidgenössische Maturitätsprüfung	
Einzelschule			
Klasse	Jahresabschlussprüfung	Jahresabschlussprüfung	Bildungsstandards, Klassenscockpit, Check 5, Orientierungsarbeit
Schülerinnen Schüler		Maturaprüfung, Basis-Check, Multi-Check, Abschlusszertifikat Volksschule	Bildungsstandards Klassenscockpit, Check 5, Stellwerk, Orientierungsarbeit

Funktionen und Bereiche von Schulleistungstests

zentriert sich der Unterricht auf assoziatives Lernen. Der Mix der Funktionen verändert grundsätzlich den Lehr-Lernprozess und damit den Unterrichtsstil. Das Ganze kann durch zusätzliche Fremdevaluationen im Rahmen des Qualitätsmanagements noch verschärft werden.

Die Abbildung gibt noch einen bedeutsamen Hinweis, der wissenschaftlich in der letzten Zeit gut untersucht wurde. Die entscheidende Einflussgrösse für den Lernerfolg ist nicht die Einzelschule, sondern die Klasse. So wurde in mehreren Untersuchungen die Auswirkung verschiedener Einflussgrössen auf die Schulleistung wie beispielsweise den Leistungsstand der Schüler, die Unterrichtsqualität, den Zeitaufwand, die Lehrmittel, das Schulklima, die Schulführung, das Qualitätsmanagement untersucht. Zentral war die Frage, welche Faktoren zur Erklärung der Schulleistung am bedeutsamsten sind. Die Ergebnisse lassen sich verkürzt in der Reihenfolge der Bedeutung wie folgt zusammenfassen:

1. Leistungsunterschiede zwischen den Schülerinnen und Schülern resultieren aus den Differenzen, welche die Lernenden bereits in die Schule mitbringen.

2. Der Einfluss der einzelnen Klasse auf die Schulleistung ist bedeutend grösser als derjenige der Einzelschule.
3. In der Klasse kommt der Lehrperson bzw. ihrem Unterricht die grösste Bedeutung im Hinblick auf die Schulleistungen der Schülerinnen und Schüler zu. Der Unterricht der Lehrperson ist somit der zentrale Schlüsselfaktor für die Schulleistung und damit auch für den Schulerfolg.

Als Folge dieser Erkenntnis konzentriert sich der Artikel auf die Leistungstests zur Unterrichtsentwicklung und individuellen Förderung.

Leistungstest zur Unterrichtsentwicklung und individuellen Lernförderung

Einige Kantone haben Schulleistungstests für klassenübergreifende Vergleiche erstellt. Mit diesen Leistungstests kann festgestellt werden, was Lernende zu einem bestimmten Zeitpunkt wissen bzw. können. Diese Standortbestimmung der Schülerinnen und Schüler gibt der Lehrperson Aufschluss darüber, wie gut die Schülerinnen und Schüler durch den

Unterricht qualifiziert werden. Diese Tests bilden eine Grundlage für die gezielte Förderung der Schülerinnen und Schüler. Sie dienen der Weiterentwicklung des Unterrichts und der Ausrichtung der individuellen Förderung. Oft fliessen sie auch ein in Selektionsentscheide und sind auch ein Indikator für die Steuerung bzw. die Wirksamkeit des Bildungssystems.

Unzweifelhaft können Leistungstests wichtige Instrumente zur Sicherung und Weiterentwicklung von Schule und damit der Unterrichtsqualität sein. Sie sind Teil eines modernen, schulischen Qualitätsmanagements in der Hand der Lehrperson. Sie verbessern die Bildungschancen jedes einzelnen Kindes, insbesondere auch von Kindern und Jugendlichen aus bildungsfernen Schichten. Zugleich bilden sie auch einen Ausgangspunkt für die gezielte Weiterentwicklung des Unterrichts und des Systems insgesamt. Als Beispiel vergleichender Tests sollen hier Klassenscockpit und Check 5 erwähnt werden. Klassenscockpit (www.klassenscockpit.ch) wie Check 5 sind Schulleistungstests zur Qualitätssicherung auf der Primarstufe. Die einzelnen Module, die jeweils Lernziele aus den Fachbereichen Deutsch und Mathematik sowie überfachliche Kompetenzen überprüfen, ermöglichen den Lehrpersonen, ihre Klasse mit einer repräsentativen Stichprobe von Schülerinnen und Schülern zu vergleichen.

Beide Testverfahren wurden wissenschaftlich entwickelt, aufgebaut und betreut. Die jeweiligen Module werden an 500 bis 600 Schülern geeicht und getestet. Die Testrückmeldung gibt der Lehrperson Auskunft über die Leistung der Klasse, einen Vergleich mit den Werten anderer Klassen (Vergleichsgruppe) und die individuellen Testergebnisse der einzelnen Schülerinnen und Schüler. Die Aufgaben der Testmodule werden laufend überarbeitet. Die Auswertung erfolgt teilweise per Internet und orientiert sich an den Zielsetzungen und den Inhalten der Lehrpläne der Kantone. Die Test-

ergebnisse erlauben teilweise auch einen Vergleich mit der eigenen Beurteilungspraxis. Hinweise zur Sorgfaltspflicht im Umgang mit den Leistungstests sollen dazu beitragen, dass die Tests nicht missbräuchlich zur Selektion oder als Kontrollinstrumente missbraucht werden. Klassenscockpit wird in mehreren Kantonen der Deutschschweiz teils obligatorisch, teils freiwillig eingesetzt.

Vom Nutzen der Leistungstests

«Das Schwein wird nicht fetter, wenn man es dauernd misst», sagt das Bonmot. Leistungstests zur formativen Diagnose haben nur dann eine Berechtigung, wenn sie zur Optimierung des Unterrichts und damit zur individuellen Förderung eingesetzt werden. Zur Verwendung und zu den Auswirkungen von Schulleistungstests auf den Unterricht liegen kaum aussagekräftige Forschungsergebnisse vor. Zur Nutzung des Leistungstests Klassenscockpit liegen zwei Befragungen von Lehrpersonen vor. Die Aussagen der Lehrpersonen in beiden Studien zeigen in der Tendenz folgendes Bild: Neugierde stand bei den Lehrpersonen an erster Stelle, gefolgt von der Standortbestimmung der Klasse im Vergleich mit anderen Klassen (90 Prozent). Rund zwei Drittel der Lehrpersonen sehen in Klassenscockpit die Möglichkeit zur Überprüfung des eigenen Beurteilungsmassstabs und zur Optimierung des Unterrichts. Nur 40 Prozent der Lehrpersonen sagen, dass sie Leistungstests zur individuellen Förderung der Schülerinnen und Schüler einsetzen. Das Ergebnis bei Klassenscockpit entspricht in etwa der Befragung von Tresch (2007) zu Check 5.

Wenn wir davon ausgehen, dass den Äusserungen der Lehrpersonen nicht immer Taten folgen, kann das Ergebnis im Vergleich zum Aufwand, den Kosten und den Nebenwirkungen nicht befriedigen. Um die Wirkung im Sinne der Optimierung des Unterrichts und

der individuellen Förderung zu erhöhen, sind folgende Massnahmen erforderlich:

- **Berechtigung:** Der Einsatz der Leistungstests durch eine Lehrperson darf erst dann erfolgen, wenn sich die Lehrperson über das notwendige Wissen ausweist. Dazu zählen testtheoretisches und statistisches Grundwissen, diagnostische Kenntnisse und Hinweise für den Umgang mit den Testergebnissen.
- **Unterstützungsangebot im Lehrerteam:** Für die gängigen Leistungstests werden von den Projektteams bereits Beratung (Hotline) und Weiterbildung angeboten. Das reicht nicht. Sinnvoll wäre es, wenn in jeder Schuleinheit eine verantwortliche Person diese Unterstützung in Form von Beratung an den konkreten Ergebnissen anbieten würde.

Unerwünschte Nebeneffekte der Leistungstests

Was für die Lehrperson zur Entwicklung des eigenen Unterrichts und zur Förderung des Lernens der Schülerinnen und Schüler gedacht ist, erweist sich als sehr anspruchsvoll und zeigt einige unerwünschte Nebeneffekte. Exemplarisch soll auf einige kritische Punkte hingewiesen werden. Selbstverständlich ist die Lehrperson für den Einsatz der Tests und der Folgen verantwortlich. Doch zeigt sich, dass in Konfliktsituationen mit Schulbehörden und Eltern die Leistungstests missbraucht werden:

- **Leistungstests als heimliche Lehrpläne?** Obwohl sich die Tests an den Lehrplänen orientieren, kommt ihnen eine normierende Funktion zu. Ein Beispiel: Die Lehrpläne des Kantons St. Gallen definieren Stufenziele für drei Jahre. Klassenscockpit gibt jedoch drei Messzeitpunkte pro Jahr vor und unterteilt damit die Stufenziele des Lehrplans in eine zeitliche und inhaltliche Vorgabe. Wurden Lehrpläne durch den



Gesetzgeber verabschiedet, so ist das bei den Testaufgaben nicht der Fall. Was im Lehrplan als Zeitraum für die individuelle Lernförderung gedacht war, wird vom Leistungstest unterlaufen.

□ **Messen die Leistungstests wirklich, was sie vorgeben, zu messen?**

Aus testtheoretischer Sicht sind die Leistungstests nicht unbedenklich. Neben der Objektivität und der Zuverlässigkeit (Reliabilität) sollten die Tests valide sein, d. h., das messen, was sie vorgeben, zu messen. Bei vielen Leistungstests ist nämlich nicht klar, welche Fähigkeiten die Testaufgaben wirklich messen. Es ist relativ einfach, durch erfahrene Lehrpersonen originelle Aufgabensammlungen zu einem Lehrplanbereich erstellen zu lassen und diese dann einer Vergleichsstichprobe vorzulegen und in Schwierigkeitsniveaus einzuteilen und ihre Trennschärfe zu bestimmen. Validität erfordert, dass für jeden einzelnen Leistungsbereich das Konstrukt geklärt und dieses mit unterschiedlichen Aufgaben überprüft wird. Dieses anspruchsvolle Vorgehen wird bei den meisten Leistungstests nicht angewandt.

□ **Stimmen Aufwand und Ertrag?**

Evaluationen zur Erfassung des Nutzens der Testinstrumente haben gezeigt, dass die Lehrpersonen die Leistungstests nur rudimentär zur Diagnose und zur Optimierung des Unterrichts verwenden. Lehrpersonen führen diese Tests durch, vergewissern sich selber, dass sie in etwa in der Streubreite der Vorgaben liegen und unterrichten wie bisher weiter. Die Chance, aufgrund der Diagnose den Unterricht für die Klasse und die einzelnen Lernenden anzupassen, wird kaum genutzt. Das Missverhältnis wird durch ein Obligatorium der Leistungstests noch verstärkt.

□ **Und letztlich: School as a place to grow up?**

Die Output-Orientierung von Schule und damit die Schulleistungstests reduzieren

den Auftrag von Schule auf einige wenige scheinbar einfach messbare und daher nützliche Bereiche. Doch Nutzen für wen und wozu? Man muss nicht das humanistische Bildungsideal bemühen, um kritisch die Entwicklung zu verfolgen. Hartmut von Hentig hat in Anlehnung an Paul Goodman gemeint, die Schule sei für Kinder und Jugendliche «a place to grow up». Schule als Ort, wo Kinder beim Hineinwachsen in eine Kultur und Gesellschaft unterstützt werden. Dieses Verständnis von Schule lässt sich nicht im kantonal genormten Testgleichschritt erreichen.

Ausgangsfrage: Verbessern Leistungstests die Schulqualität?

Im engen Sinne lautet die Antwort NEIN. Leistungstests machen Aussagen über die Leistungen der Schülerinnen und Schüler einer Klasse und damit über deren Unterricht. Der Einfluss der einzelnen Klasse auf die Schulleistung ist bedeutend grösser als derjenige der einzelnen Schule. Unter den folgenden Voraussetzungen und Bedingungen können Leistungstests zur Optimierung des Unterrichts und zur individuellen Förderung der Schülerinnen und Schüler beitragen:

- Leistungstests unterliegen dem Primat der Lehrpläne.
- Die Schulbehörden fokussieren die Funktion der Leistungstests primär im Sinne der Unterrichtsentwicklung und regeln die Verbindlichkeit und Kommunikation.
- Die Rückmeldungen der Testergebnisse an die Lehrpersonen sind verständlich und aussagekräftig.
- Die Lehrpersonen verfügen über die notwendigen Grundkenntnisse der Testtheorie und der Statistik zur Analyse der Testergebnisse.
- Die Lehrpersonen können aufgrund der Analyse der Testergebnisse die Leistungen

der Lernenden diagnostizieren und adaptiv didaktische Massnahmen für die Steigerung der Unterrichtsqualität und die individuelle Förderung der Lernenden planen und umsetzen.

- Die Lehrpersonen werden bei der Umsetzung und Überprüfung der geplanten Massnahmen im Rahmen der Unterrichtsentwicklung vor Ort (Schulleitung, qualifizierte Kolleginnen und Kollegen) unterstützt.

Mit Tests verhält es sich wie mit anderen Werkzeugen: Sie erfordern einen verantwortungsvollen und zweckbezogenen Einsatz. Wer nur einen Hammer als Werkzeug besitzt, dem besteht die Welt nur aus Nägeln.

Literaturhinweise:

- Helmke, A. (2003). Unterrichtsqualität. Erfassen, Bewerten, Verbessern. Kallmeyer: Seelze.
- Tresch, S. (2007). Potential Leistungstest. Wie Lehrerinnen und Lehrer Ergebnisrückmeldungen zur Sicherung und Steigerung ihrer Unterrichtsqualität nutzen. h.e.p.: Bern.



Zum Autor:
Prof. Dr. Titus Guldemann
 Prorektor Forschung, Entwicklung und Beratung der Pädagogischen Hochschule des Kantons St. Gallen
 Spezialgebiete: Lehrerbildung, Lehr-Lernforschung, Schul- und Organisationsentwicklung, Schulevaluation